

Université de Montréal

**Étude de la validité d'un instrument de mesure
de la compétence informationnelle
: l'exemple du QuizCI**

par Catherine Séguin

Département d'administration et fondement de l'éducation
Faculté des sciences de l'éducation

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de Maîtrise en mesure et évaluation en éducation

Août 2016

© Catherine Séguin, 2016

Résumé

Cette recherche documente la validité du QuizCI, un instrument pour mesurer la compétence informationnelle (CI). Le QuizCI a été développé à l'Université du Québec en Outaouais, une université francophone du Canada. L'instrument est composé de 32 items à choix de réponses.

La preuve de validité est documentée selon les types de validité proposés par Laveault et Grégoire (2002). Pour supporter l'élaboration du QuizCI, des arguments de validité apparente et de contenu ont été colligés. Avec l'instrument élaboré, un échantillon de données issues de 469 étudiants a été collecté de 2012 à 2015. Cet échantillon a été utilisé pour réaliser les analyses produisant l'information destinée à documenter la validité conceptuelle du QuizCI. Parmi les informations prises en compte se trouvent l'indice de simplicité, l'indice de discrimination, le coefficient alpha de Cronbach, la solution de l'analyse factorielle exploratoire et l'information disponible après l'application d'un modèle de Rasch.

L'étude permet de conclure que le QuizCI mesure la CI avec une certaine validité, mais fait aussi le constat de différents problèmes affectant la mesure. Certains items sont problématiques. Puis, l'instrument dans son ensemble est facile pour les étudiants, ce qui nuit à la validité de la mesure. Ainsi, ce premier portrait en regard de la validité du QuizCI constitue un point de départ pour son amélioration.

Enfin, peu d'auteurs présentent l'étude de la validité d'un instrument mesurant la compétence informationnelle, de surcroît en contexte francophone. Aussi, la présente étude pourra fournir un exemple de l'usage de ces méthodes.

Mots-clés : compétence informationnelle, mesure, validité, validation, instrument, test, indice de simplicité, indice de discrimination, cohérence interne, analyse factorielle exploratoire, modèle de Rasch

Abstract

This research documents the validity of QuizCI, an instrument for measuring information literacy (CI). The QuizCI was developed at the Université du Québec en Outaouais, a French-language university in Canada. The instrument has 32 multiple choice items.

Evidence of validity is documented according to the types of validity Laveault and Gregory (2002). To support the development of QuizCI, arguments of apparent and content validity were collected. With the instrument developed, a sample of data from 469 students was collected from 2012 to 2015. This sample was used to perform analyzes and produce information intended to document the conceptual validity of QuizCI. Among the information taken into account are the simplicity of index, the discrimination, the Cronbach's alpha, the exploratory factor analysis solution available information after the application of a Rasch.

After which, the study concludes that the QuizCI measurement IC with some validity, but also the finding various problems affecting the measurement. Some items are problematic. Then, the whole instrument is easy for students, which affects the measurement. Thus, this first portrait of the validity of QuizCI is a starting point for improvement.

Finally, not many authors present study of the validity of an instrument to measure Information Literacy, especially in French-speaking context. Also, this study may provide an example to use of these methods.

Keywords : information literacy, measurement, validity, validation, instrument, test, index of simplicity, discrimination index, internal consistency, exploratory factor analysis, Rasch

Table des matières

| | |
|--|------|
| Résumé | ii |
| Abstract..... | iii |
| Liste des tableaux | vii |
| Liste des figures..... | viii |
| Liste des sigles..... | ix |
| Liste des abréviations..... | xi |
| Remerciements | xiii |
| Introduction..... | 1 |
| Chapitre 1 : problématique | 3 |
| 1.1. La compétence informationnelle..... | 3 |
| 1.2. Former à l'usage de la compétence informationnelle | 7 |
| 1.3. Question générale de recherche | 10 |
| Chapitre 2 : cadre conceptuel et recension des écrits | 13 |
| 2.1. Cadre conceptuel de la validité | 13 |
| 2.1.1. Les types de validités | 14 |
| 2.2. Recension des écrits étudiant la validité | 19 |
| 2.2.1. Les expérimentations québécoises | 20 |
| 2.2.2. Les études complémentaires | 23 |
| 2.3. Recension des dispositifs pour colliger la preuve de validité | 25 |
| 2.3.1. Dispositifs pour constituer la preuve de validité apparente | 25 |
| 2.3.2. Dispositifs pour constituer la preuve de validité de contenu | 28 |
| 2.3.3. Dispositifs pour constituer la preuve de validité en référence à un critère | 30 |
| 2.3.4. Dispositifs pour constituer la preuve de validité conceptuelle | 31 |
| 2.4. Buts de la présente étude..... | 39 |
| 2.4.1. Validité apparente et de contenu | 39 |
| 2.4.2. Validité en référence à un critère | 40 |
| 2.4.3. Validité conceptuelle | 40 |
| 2.5. Questions spécifiques de recherche | 41 |
| Chapitre 3 : cadre méthodologique | 42 |

| | |
|---|----|
| 3.1. Type de recherche et justifications | 42 |
| 3.2. Visées de la recherche..... | 42 |
| 3.3. Élaboration du QuizCI | 42 |
| 3.3.1 Validité de contenu (interne) | 43 |
| 3.3.2 Validité apparente | 46 |
| 3.3.3 Validité de contenu (externe)..... | 46 |
| 3.3.4. Résultats sommaires du processus de validation préalable du QuizCI..... | 49 |
| 3.4. Participants..... | 51 |
| 3.5. Les données..... | 53 |
| 3.6. Méthodologies pour documenter la validité conceptuelle | 53 |
| Chapitre 4 : résultats | 55 |
| 4.1. Analyse descriptive des données | 55 |
| 4.1.1 Distribution des scores totaux | 55 |
| 4.1.2. Taille d'échantillon | 56 |
| 4.1.3. Homogénéité de la variance..... | 57 |
| 4.2. Analyses et résultats issus de la théorie classique des tests | 59 |
| 4.2.1. Difficulté de l'item..... | 59 |
| 4.2.2. Discriminer les étudiants en fonction de leur CI | 60 |
| 4.2.3. Cohérence interne | 62 |
| 4.3. Analyses et résultats pour documenter le concept de CI | 65 |
| 4.3.1. Analyses préalables..... | 65 |
| 4.3.2. Analyse factorielle exploratoire | 66 |
| 4.4. Application du modèle de Rasch et informations sur la validité | 72 |
| 4.4.1. Postulats de la théorie de réponse à l'item..... | 73 |
| 4.4.2. Processus d'ajustement des données au modèle de Rasch..... | 75 |
| Chapitre 5 : discussion sur la validité du QuizCI | 86 |
| 5.1. Est-ce que le QuizCI réagit en fonction de la mesure captée?..... | 86 |
| 5.2. Est-ce que le QuizCI représente la CI dans sa nature et sa structure conceptuelle? | 87 |
| 5.3. Est-ce que les données collectées avec le QuizCI constituent une mesure de la CI? | 89 |
| 5.4. Décisions pour l'amélioration de la contribution de l'item à la mesure | 89 |
| 5.3. Limites de l'étude | 94 |

| | |
|---|----|
| 5.4. Prospectives | 95 |
| Conclusion | 96 |
| Les sources documentaires | 97 |
| Annexe 1 : Évaluation de la formation par les étudiant..... | I |

Liste des tableaux

| | |
|--|-----------|
| Tableau 1. Normes pour opérationnaliser la maîtrise de l'information | p. 5, 6 |
| Tableau 2. Activités de formation documentaire à l'UQO – St.Jérôme (FD.UQO.SJ) ... | p. 9 |
| Tableau 3. Reproduction du tableau « Types de validité » | p. 14 |
| Tableau 4. Documents québécois recensés | p. 21 |
| Tableau 5. Cinq études complémentaires recensées | p. 23 |
| Tableau 6. Règle de décision de l'indice de simplicité (IS) | p. 33 |
| Tableau 7. Règle de décision de l'index de discrimination (d) | p. 34 |
| Tableau 8. Tableau de spécification du QuizCI | p. 45 |
| Tableau 9. Processus pour documenter la validité apparente et de contenu | p. 48, 49 |
| Tableau 10. Données collectées et échantillon retenus pour les analyses ultérieures..... | p. 52 |
| Tableau 11. Indice de simplicité des items du QuizCI. | p. 60 |
| Tableau 12. Capacité de discrimination de l'item du QuizCI | p. 62 |
| Tableau 13. Synthèse des informations sur la validité du QuizCI. | p.64 |
| Tableau 14. Qualité de représentation des facteurs par les données de l'échantillon ... | p. 68 |
| Tableau 15. Solution factorielle retenue | p. 69, 70 |
| Tableau 16. Étiquettes de la solution factorielle des items du QuizCI. | p. 71 |
| Tableau 17. Effets du processus d'ajustement sur les paramètres des items | p. 79 |
| Tableau 18. Effets du processus d'ajustement sur les paramètres des items..... | p. 79 |
| Tableau 19. Mesures de la difficulté des items du QuizCI sur l'échelle de la CI | p. 84 |
| Tableau 20. Informations sur la validité de la mesure QuizCI et décisions pour l'amélioration du QuizCI | p. 92. |

Liste des figures

| | |
|---|-------|
| Figure 1. Facteurs influençant l'étude de validation apparente. | p. 26 |
| Figure 2. Histogramme des scores totaux du QuizCI. | p. 56 |
| Figure 3. Dispersion des scores des participants | p. 57 |
| Figure 4. Moyenne des groupes avec barres d'erreur | p. 58 |
| Figure 5. Tracé d'effondrement des valeurs propres des données collectées avec le QuizCI. | p. 67 |
| Figure 6. Coefficient Infit des items du QuizCI | p. 76 |
| Figure 7. Cartes des personnes et des items. | p. 83 |

Liste des sigles

| | |
|-------------|---|
| ACRL | Association of College and Research Libraries |
| AFE | Analyse factorielle exploratoire |
| ALA | American Libraries Association |
| ANZIL | Australian and New Zealand Information Literacy Framework |
| ANOVA | Analyse de variance univariée |
| CI | Compétence informationnelle |
| CM | Carré moyen |
| CRÉPUQ | Conférence des recteurs et des principaux des universités du Québec |
| Ddl | Degré de liberté |
| EDC | Evidence-Centered-Design |
| EFE | Évaluation de la formation par les étudiants |
| EEE | Évaluation de l'enseignement par les étudiants |
| ENA | Environnement numérique d'apprentissage |
| FD | Formation documentaire |
| FD-UQO-SJ. | Formation documentaire à l'Université du Québec en Outaouais- campus Saint-Jérôme |
| GPA | Grade Point Average |
| GPCM | General Partial Credit Model |
| ICT | Information and communication technologies |
| IL | Information literacy |
| ISST | Information Seeking Skills Test |
| KMO | Kaiser-Meyer-Olkin |
| ORL | Online Research Lab |
| r* | Activité récurrente à chaque formation documentaire |
| SAILS | Standards Assess of Information Literacy Skills |
| SCONUL | Society of College, National and University Libraries |
| SOI | Sciences infirmières |
| TECFÉE | Tests de certification en français écrit pour l'enseignement |
| TCT | Théorie classique des tests |

TIC Technologies de l'information et de la communication

TRI Théorie de réponse à l'item

TSO Travail social

VA Validité apparente

VC Validité de contenu

Zstd Standardisé

Liste des abréviations

sig..... Significatif

corr..... Corrélation

À ma famille, pour votre soutien dans cette belle aventure!

Remerciements

Je voudrais exprimer toute ma reconnaissance à ma Directrice de recherche Nathalie Loye, pour son accompagnement et son soutien tout au long du parcours.

Mes sincères remerciements aux membres du jury Sébastien Béland, Jean-Guy Blais, puis Roseline Garon. Leurs rétroactions ont contribué à l'amélioration de ce travail.

Mes sincères remerciements à tout le personnel du Département d'administration et fondement en éducation.

Introduction

À une époque où la documentation abonde plus qu'à n'importe quelle autre, la capacité à utiliser l'information pour répondre efficacement aux exigences du quotidien constitue une plus-value. La société de l'information¹ met à l'épreuve la capacité des individus à exploiter l'information afin d'améliorer leur vie « tant scientifique, citoyenne, démocratique, laïque, sociale, économique, culturelle, humaniste, pédagogique et cognitive » (Groupe de Recherche sur la Culture et la Didactique de l'Information, 2010, p. 661). Ainsi, les habiletés à localiser, accéder à et évaluer l'information sont essentielles à l'individu pour combler ses besoins, adopter un regard critique sur son environnement et faire des choix éclairés. L'information permet par exemple d'être un consommateur plus averti, un patient plus responsable ou un citoyen plus impliqué. Puisqu'elle a un impact direct et fondamental sur la capacité à améliorer les conditions de vie, l'habileté à la *recherche d'information*² a été décrétée « condition de vie dans la société de l'information » par l'Organisation des Nations Unies pour l'éducation, la science et la culture (2003). La littérature sur le sujet s'avère unanime sur les bénéfices permis par la maîtrise de l'information, de même que sur la nécessité d'agir pour la favoriser. Les observateurs mondiaux en appellent même à un mouvement social et politique pour l'équité des chances par l'éducation à la « maîtrise de l'information » (International Federation of Library Associations, 2011). Bref, le droit à l'information doit s'opérationnaliser par l'entremise de l'accès aux habiletés nécessaires à son usage.

En milieu universitaire, l'information constitue la matière première des activités, tant pour former des diplômés, que pour être concurrentiel en matière de recherche et de développement. Le bibliothécaire universitaire a pour fonction de favoriser l'accès à l'information dans un contexte, d'une part, marqué par l'abondance documentaire, d'autre part, où l'environnement informationnel s'est complexifié par la variété des supports et des formats, alors que la démocratisation de l'enseignement supérieur élargit constamment la communauté d'étudiants à desservir. Pour favoriser l'accès à l'information de façon autonome, efficace et au

¹L'expression « société de l'information » sera utilisée dans ce texte comme équivalent de « société du savoir ».

²L'expression « recherche d'information » sera utilisée de façon quasi équivalente à « recherche documentaire », bien qu'il faut trouver le document pour y chercher l'information.

plus grand nombre, le bibliothécaire universitaire propose des activités de formation pour maîtriser l'information.

Cette recherche porte sur les habiletés permettant la maîtrise de l'information et plus précisément l'objet d'apprentissage qui permet d'atteindre cet objectif, la compétence informationnelle (CI). Ce premier chapitre résume les fondements théoriques encore mouvants de la CI. Puis, il résume l'initiative de formation documentaire à l'origine de cette étude, soulevant des problématiques liées à la qualité de la mesure de la CI.

Le second chapitre expose d'une part le cadre conceptuel. Il décrit la typologie de Laveault et Grégoire (2002) utilisée dans la présente étude pour catégoriser la documentation sur la validité du QuizCI, ainsi que l'apport des domaines de la validité décrits par Borsboom, Mellenbergh et van Heerden (2004) à la preuve de validité conceptuel. D'autre part, le second chapitre présente la recension d'écrits documentant la validité de leur instrument afin d'en relever les dispositifs et méthodologies employés à cette fin.

Le troisième chapitre précise comment l'instrument QuizCI a été élaboré. Il relate de quelle façon il a été administré. Il décrit ensuite les données ainsi collectées. Puis, il annonce les analyses retenues pour produire de l'information la validité de l'instrument et des données.

Enfin, le quatrième expose le résultat des analyses réalisées. Alors que le dernier chapitre discute de l'apport de ces résultats à documenter la validité du QuizCI et des données qu'il collecte à représenter la CI des étudiants.

Chapitre 1 : problématique

1.1. La compétence informationnelle

La littérature anglophone a résumé le champ de compétence ici visé par l'expression *Information Literacy* (Zurkowski, 1974). L'Organisation des Nations unies pour l'éducation, la science et la culture (UNESCO) l'a traduit en français par l'expression *maîtrise de l'information* (Catts et Lau, 2008). Cette locution résume l'objectif visé. La *maîtrise de l'information* a été définie comme un « ensemble de compétences permettant de reconnaître l'existence d'un besoin d'information, d'identifier l'information adéquate, de la trouver, de l'évaluer et de l'exploiter en relation avec une situation donnée, dans une perspective de résolution de problème » (Office national de la langue française, 2008).

Pour opérationnaliser la maîtrise de l'information en tant qu'objet d'apprentissage, l'Association of College and Research Libraries (ACRL) a publié en 2000 les « Information Literacy Competency Standards for Higher Education ». À leur tour, différents organismes phares en sciences de l'information ont proposé des normes [traduction de *standards*] pour opérationnaliser la maîtrise de l'information. Hinchliffe (2005) a produit une synthèse non exhaustive de normes visant à orienter les pratiques pour la maîtrise de l'information. L'auteure met en perspective différentes normes faisant autorité dans diverses régions du monde : dans l'est des États-Unis, l'Association of College and Research Libraries (2000), division de l'American Libraries Association (ALA); dans l'ouest des États-Unis à la Washington University (Seattle, WA), l'Information & Technologies skills for success (Big6); au Royaume-Uni la Society of College, National and University Libraries (SCONUL); ou encore en Océanie l'Australian and New Zealand Information Literacy Framework (ANZIL).

La norme de l'ACRL fait autorité au Québec. Elle énonce les manifestations observables pour reconnaître *un usager compétent avec l'information* [traduction libre] et fait l'objet du tableau 1, colonne 1. La norme ACRL a été traduite en français pour orienter les pratiques dans les universités québécoises par la Conférence des recteurs et des principaux des universités québécoises. Ce document a été intitulé : « Compétences informationnelles : niveau recommandé à l'entrée au 1er cycle universitaire » (Conférence des recteurs et des principaux des universités du Québec, 2005) [tableau 1, col. 2]. Depuis, au Québec, cet objet d'apprentissage est souvent nommé

compétence informationnelle (CI). La CI a été synthétisée l'année suivante en 4 étapes par le réseau de l'Université du Québec (2006) [tableau 1, col. 3]. Les établissements universitaires québécois adhèrent assez librement à la norme, comme en témoigne l'hétérogénéité des formulations sur leur page Web, rapportée aux colonnes 4 à 9 du tableau 1.

Le Ministère de l'Éducation, du Sport et du Loisir a aussi reconnu l'importance de la maîtrise de l'information particulièrement lors de l'usage des technologies de l'information et de la communication (TIC). Au sein du *Programme de l'école québécoise aux niveaux primaire et secondaire* (Ministère de l'Éducation du Sport et du Loisir, 2006), il fournit l'énoncé de la *Compétence TIC*. Elle est ensuite appliquée au niveau post-secondaire dans le cadre du *Profil TIC au collégial* produit par la Fédération des Cégeps (Réseau des répondantes et répondants en technologies de l'information et de la communication, 2011). Ces énoncés figurent aux colonnes 10 et 11 du tableau 1. La CI et la compétence TIC se recoupent.

Encore en 2013, l'UNESCO souligne la difficulté de circonscrire la CI (2013) et divers auteurs déplorent le manque de données permettant d'en formaliser le construit (Catts et Lau, 2008; Groupe de Recherche sur la Culture et la Didactique de l'Information, 2010).

En somme le concept de CI, objet de la présente recherche, n'est pas conceptuellement fixé.

Tableau 1

Normes pour opérationnaliser la maîtrise de l'information.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|--|---|---|---|--|---|---|---|---|---|--|
| Norme ACRL (Traduction de), 2000 | Conférence des recteurs et principaux des universités du Québec, 2005 | Université du Québec, 2006 | Université du Québec à Montréal, 2010 | Université Laval, 2011 | Université du Québec à Rimouski, 2010 | Université de Montréal, 2002 | École des technologies supérieures, 2011 | Université du Québec en Abitibi- Témiscamingue, 2007 | Compétence TIC, 2008 | Profil TIC au collégial, 2011 |
| Détermine la nature et l'étendue de l'information dont il a besoin. | Préciser le besoin d'information | Préparer sa recherche | La personne compétente dans l'usage de l'information reconnaît son besoin d'information et sait déterminer la nature et l'étendue de l'information nécessaire pour y répondre. | L'étudiant qui possède les compétences informationnelles détermine la nature et l'étendue de l'information dont il a besoin. | Phase de préparation de la recherche | Reconnaît un besoin d'information et détermine son nature et son étendue; | Clarifier des concepts | Deuxième intervention : préciser son sujet de recherche | | |
| Accède avec efficacité et efficience à l'information dont il a besoin. | Accéder à l'information | Repérer l'information Obtenir les documents | La personne compétente dans l'usage de l'information sait accéder avec efficacité et efficience à l'information dont elle a besoin. | L'étudiant qui possède les compétences informationnelles accède avec efficacité et efficience à l'information dont il a besoin. | Phase de recherche | Accède avec efficacité à l'information dont il a besoin; | Repérer l'information | Troisième intervention : bâtir une stratégie de <u>recherche</u> Quatrième intervention : la recherche documentaire | Exploiter les TIC | Maîtriser l'environnement <u>de travail</u> Rechercher de <u>l'information</u> Communiquer et collaborer sur Internet |
| Évalue de façon critique tant l'information que ses sources | Évaluer l'information | Analyser la documentation | La personne compétente dans l'usage de l'information sait évaluer de façon critique tant l'information que les sources dont elle est tirée et sait intégrer cette information à ses connaissances personnelles et à son système de valeurs. | L'étudiant qui possède les compétences informationnelles évalue de façon critique tant l'information que ses sources et intègre l'information sélectionnée à sa base de connaissances personnelle et à son système de valeurs. | Phase d'analyse | Évalue de façon critique tant l'information que les sources d'information; | Évaluer l'information | Cinquième intervention : l'évaluation des sources | Évaluer l'efficacité de l'utilisation de la technologie. | Traiter l'information Évaluer le projet |

(Tableau se continu)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|--|----------------------------------|---------------------|---|---|--------------------|---|--|---|---|-------------------------|
| Intègre l'information sélectionnée à sa base de connaissances personnelle et à son système de valeurs. | | | La personne compétente dans l'usage de l'information reconnaît l'importance d'acquérir des compétences informationnelles dans la perspective d'une formation continue. | | | | | | | |
| Utilise efficacement l'information, individuellement ou comme membre d'un groupe, en vue d'atteindre un objectif spécifique. | Exploiter l'information | Rédiger son travail | La personne compétente dans l'usage de l'information sait développer, individuellement ou comme membre d'un groupe, de nouvelles connaissances en intégrant l'information à ses connaissances initiales. La personne compétente dans l'usage de l'information sait utiliser l'information recueillie ou nouvellement générée pour réaliser ses travaux | L'étudiant qui possède les compétences informationnelles utilise efficacement l'information, individuellement ou comme membre d'un groupe, en vue d'atteindre un objectif spécifique. | Phase de diffusion | Utilise l'information individuellement ou comme membre d'une équipe de façon efficace; | | | Utiliser les technologies de l'information et de la communication pour effectuer une tâche. | Présenter l'information |
| Comprend plusieurs des questions économiques, juridiques et sociales relatives à l'utilisation de l'information accède à l'information et l'utilise de façon éthique et conformément à la loi. | Tenir compte des enjeux éthiques | | La personne compétente dans l'usage de l'information comprend les enjeux culturels, éthiques, légaux et sociaux liés à l'usage de l'information et se conforme aux exigences éthiques et légales liées à cet usage. | L'étudiant qui possède les compétences informationnelles comprend plusieurs des questions économiques, juridiques et sociales relatives à l'utilisation de l'information et il accède à l'information et l'utilise de façon éthique et conformément à la loi. | | comprend les enjeux éthiques, légaux et sociaux et respecte les exigences éthiques et légales reliées à l'utilisation de l'information. | Réutiliser l'information en respectant le droit d'auteur | Première intervention : les citations et le plagiat | | |

1.2. Former à l'usage de la compétence informationnelle

En milieu universitaire, de nombreux bibliothécaires ont élaboré des activités pour favoriser la maîtrise de l'information des étudiants (Service de soutien à la formation - Université de Sherbrooke, 2011; Université de Montréal, 2012; Université du Québec, 2006; Université Laval, 2012). Les activités d'enseignement de la CI sont couramment désignées par l'expression : formation documentaire (FD). Au début, les FD visaient les aspects techniques de la recherche d'information pour interroger les ressources documentaires (bases de données ou autres). Toutefois, les bibliothécaires constatent quotidiennement que le savoir procédural seul ne permet pas de rendre l'étudiant autonome ou efficace dans ses recherches d'information. Pour réaliser son travail universitaire, l'étudiant doit pouvoir énoncer son besoin d'information et évaluer de quelle façon la documentation repérée contribue à y répondre. Une fois l'information sélectionnée, l'étudiant doit pouvoir la consigner, puis l'utiliser de façon efficace et éthique. Ainsi, la compétence informationnelle s'exprime au cours du processus où l'étudiant passe de l'identification de son besoin documentaire, à sa satisfaction.

L'Université du Québec en Outaouais a développé en 2006 le service de bibliothèque sur le campus à Saint-Jérôme. Dès lors, une formation documentaire (FD) a été élaborée, pour les étudiants inscrits au programme de baccalauréat en sciences infirmières et au programme de baccalauréat en travail social. Une séance de FD durait 1 h 00 et était intégrée à un cours obligatoire, typiquement suivi à la première session universitaire. Il est rapidement apparu que ce format était insuffisant pour rendre les étudiants autonomes.

Pour mieux intégrer ce contenu et optimiser le temps de cours alloué à la FD, un travail d'arrimage a été entrepris entre la bibliothécaire (auteure de la présente étude) et le professeur disciplinaire. Au cours de cet exercice, une analyse plus fine des contenus a permis de constater qu'en fait, cet objet d'apprentissage était déjà présent tant parmi les objectifs de cours, que parmi les objectifs de programme, ainsi que dans les critères réguliers de correction pour les travaux universitaires (Grenier & Ouellette, 2012; Verdon, 2012). Ainsi, les composantes de la CI sous-tendent directement la qualité des travaux universitaires, particulièrement la qualité des sources

utilisées et l'application de la forme prescrite des travaux pour faire un usage éthique de l'information.

Ce constat a contribué à l'allocation de plus de temps de cours aux activités de FD et à l'élaboration d'une variété d'activités spécifiquement dédiées à cet apprentissage. La FD a alors pris une nouvelle ampleur. Elle a été scindée en plusieurs séances pour offrir aux étudiants différentes opportunités de s'approprier les éléments de cet objet d'apprentissage, de travailler sur des activités d'apprentissage et de recevoir des rétroactions. Les activités de la *formation documentaire de l'Université du Québec en Outaouais- campus Saint-Jérôme (FD-UQO-SJ)* de laquelle a émergé la présente étude sont décrites au tableau 2.

Cette formule, la FD-UQO-SJ, a prévalu de la session d'automne 2009 à la session d'hiver 2015³. À titre indicatif, à l'automne 2012 cette série d'activités a rejoint 186 étudiants, comme décrit au tableau 9 (p.41). L'objet d'apprentissage de la FD a été élaboré selon la norme ACRL (2000), traduite (2005) et révisée (2008). La définition de la CI retenue pour l'élaboration du QuizCI est celle décrite par la norme ACRL traduite et révisée (2008).

L'objectif général de la FD était de rendre l'étudiant suffisamment compétent pour réaliser de façon autonome des travaux de qualité universitaire, selon les normes établies. Pour y contribuer, des évaluations ont été intégrées à la FD-UQO-SJ. Parmi celles-ci, les activités #6 et #8, de type *résolution de problème*, sont deux moitiés d'un questionnaire visant à vérifier l'acquisition des connaissances transmises, leur compréhension et leur capacité à être utilisées dans le cadre de mises de situation plausibles. Le QuizCI⁴ est composé de 32 questions à choix multiples. Ces activités contribuent à la note du cours dans lequel la FD est intégrée; respectivement #6 (4%) et #8 (5%).

³ Suite aux résultats de la présente étude, le QuizCI a été modifié.

⁴ L'élaboration du QuizCI est décrite en section 3.4.

Tableau 2

Activités de formation documentaire à l'UQO – St-Jérôme

| # | Activités | Acteur |
|----|---|----------------|
| 1 | Accueil des étudiants, présentation de la bibliothèque, de son rôle et des personnes-ressources | Bibliothécaire |
| 2 | Présentation du plan de cours, ainsi que de la formation documentaire dans le cadre du cours et du programme | Enseignant |
| 3 | Introduction aux services de bibliothèque Introduction au service de bibliothèque : 30 minutes sur l'environnement informationnel de l'UQO. | Bibliothécaire |
| 4 | Séance 1 : prétest | Bibliothécaire |
| 5 | Formation documentaire alternant théorie, exercices pratiques, évaluations formatives et rétroactions; porte sur les dimensions 3 et 5 de la norme CRÉPUQ | Bibliothécaire |
| 6 | QuizCI partie 1 : 12 questions à répondre à la maison, seul, sur la matière vue en classe, avec consultation de l'ensemble de la documentation disponible et de l'environnement informationnelle | L'étudiant |
| 7 | Séance 2 Formation documentaire alternant théorie, exercices pratiques, évaluations formatives et rétroactions; porte sur les dimensions 1, 2 et 4 de la norme CRÉPUQ | Bibliothécaire |
| 8 | QuizCI partie 2 : 20 questions à répondre à la maison, seul, sur la matière vue en classe, avec consultation de l'ensemble de la documentation disponible et de l'environnement informationnel | L'étudiant |
| 9 | 2 travaux à remettre à l'enseignant, présumant l'étudiant compétent à la recherche et exigeant d'utiliser les notions vues durant la formation documentaire. | Enseignant |
| 10 | Séance 3 Rétroaction sur le QuizCI, explications complémentaires, opportunités de poser des questions, etc. | Bibliothécaire |

(Tableau se continu)

| # | Activités | Acteur |
|----|--|----------------|
| 11 | Liste de références Réalisation d'un travail hors cours d'intégration pour une évaluation sommative | L'étudiant |
| 12 | Évaluation de la formation par les étudiants (EFE) disponible en annexe 1. | Bibliothécaire |
| 13 | Évaluation de l'enseignement par les étudiants (EEE) dans le cadre du cours | Enseignant |

1.3. Question générale de recherche

En tant que formatrice et auteure du QuizCI, nous étions concernée par la qualité de l'outil proposé. Nous avons plusieurs interrogations auxquelles nous ne pouvions répondre avec certitude :

- Comment savoir s'il y a suffisamment de questions pour estimer le niveau d'habileté des étudiants? Comment vérifier si les items permettent de discriminer adéquatement les étudiants en fonction de leur CI?
- Comment savoir si les items permettent de bien représenter la CI, par le contenu des questions?
- Ou même, comment savoir si le niveau de difficulté des questions est adéquat pour nos étudiants?

Comme ailleurs, en milieu universitaire se trouve à la fois une concurrence féroce pour la disponibilité des ressources, ainsi que de multiples processus visant à répondre à l'imputabilité en regard de la qualité de l'enseignement offert aux étudiants. Pourtant, la formation documentaire constitue une forme d'enseignement qui n'est pas encadrée par les modes habituels de contrôle comme l'évaluation de programme ou l'évaluation de l'enseignement. Aussi, les énergies investies débouchent sur des pratiques peu encadrées ou documentées.

En effet, l'absence de données validées sur la CI des étudiants a été documentée. L'absence de données validées permet difficilement au bibliothécaire-formateur de rendre

compte de l'efficacité des actions et de l'efficience des ressources investies pour l'atteinte des objectifs de formation (DeMars, June 2-5, 2002; DeMars, Cameron et Erwin, 2003; Megan, 2009; Oakleaf, Millet et Kraus, 2011; Shipman, Aloï et Jones, 2003; Staley, Branch et Hewitt, 2010; Timmers et Glas, 2010). L'absence de données validées ne permet pas de démontrer la qualité des démarches entreprises par le bibliothécaire-formateur (Katz, 2007b; Mery, Newby et Peng, 2011) et ainsi favoriser l'imputabilité des acteurs, tel que réclamé par certains ordres professionnels ou administrations universitaires (Business-Higher Education Forum, 2003; Kent State University, 2015).

Dans le même esprit, l'absence de données validées ne permet pas au service de bibliothèque de démontrer l'apport des CI (DeMars et al., 2003) pour répondre aux politiques sociales, nationales et/ou institutionnelles pour l'équité des chances par l'éducation (Department of Education and Science (U.S.), 2002; Katz, 2007a; O'Connor, Radcliff et Gedeon, 2001; Scharf, Elliot, Huey, Briller et Joshi, 2007).

Ainsi, être capable de produire des données représentant de façon valide la CI des étudiants permettrait de mieux cibler le rôle de la formation documentaire dans le parcours de l'étudiant. Ce type de données permettrait de clarifier l'apport du bibliothécaire-formateur et d'optimiser cet apport. Puis, plus largement, de telles données permettraient de mieux démontrer le rôle et l'apport de la CI dans le parcours universitaire.

De façon similaire, pour documenter la validité de leurs données, des chercheurs ont puisé parmi des méthodologies utilisées en psychométrie pour mesurer une compétence (DeMars et al., 2003; Katz, 2007b; Mery et al., 2011; O'Connor et al., 2001; Timmers et Glas, 2010). Par exemple les auteures du reconnu test SAILS ont formulé cette quête comme documenter la validité de leur instrument et des données qu'il collecte (O'Connor et al., 2001). Dans le même but, la présente étude compte réaliser une démarche de validation pour documenter la validité du QuizCI et des données qu'il collecte.

En examinant la question de plus près, à l'instar de ces chercheurs, nous constatons que mesurer la CI consiste à mesurer une compétence, soit un trait latent indirectement observable.

En effet, une compétence s'observe difficilement. Le questionnaire est censé permettre de la représenter. Étudier sa validité consiste à vérifier s'il fait son travail.

Ainsi, la question générale de recherche de la présente étude est : comment documenter la validité du QuizCI, un instrument visant à mesurer la CI des étudiants à l'UQO Saint-Jérôme?

Cette quête de légitimité nous a conduite dans le labyrinthe de la validité. Les éléments permettant de comprendre la notion de validité font l'objet de la recension des écrits au second chapitre. Cette recension a permis de préciser les questions spécifiques de recherche de la présente étude.

Chapitre 2 : cadre conceptuel et recension des écrits

Ce second chapitre expose le cadre théorique retenu pour discuter de la validité du QuizCI, ainsi que les dispositifs mis en place par différents chercheurs pour produire l'information constituant des évidences de validité. Dans une première section, le cadre conceptuel choisi pour étudier la validité du QuizCI est décrit et expliqué. Une seconde section présente les écrits recensés étudiant la validité de leur instrument. De ces études seront relevés, d'une part, les dispositifs utilisés pour documenter la validité; d'autre part, le résultat obtenu par chacun en regard de la démonstration de validité. Au final, ce chapitre permet de cibler la méthodologie à privilégier pour documenter la validité du QuizCI.

2.1. Cadre conceptuel de la validité

Que signifie ou qu'implique *examiner la validité du QuizCI et des données qu'il collecte*? En fait, il s'agit d'étudier les caractéristiques et les qualités psychométriques des données, qui permettent de faire des inférences sur la qualité de l'instrument et de ses items à faire le travail qu'ils sont censés faire. Dans le cas du QuizCI, il s'agit de représenter la CI des étudiants.

Cureton (1951) a résumé la notion de validité comme l'étude de la qualité d'un instrument à collecter des données représentant ce qu'elles sont censées représenter. Panicker (2005) la décrit comme référant à la justesse, la pertinence et la signification des inférences possibles à partir des données collectées. Selon cette dernière conception, les données sont présumées valides lorsqu'elles permettent d'atteindre avec justesse les objectifs pour lesquels elles ont été collectées et lorsqu'elles sont pertinentes et signifiantes pour les utilisations voulues. En ce sens, certaines preuves de validité peuvent légitimer l'utilisation des données collectées, par exemple, pour rendre compte de l'efficacité des actions et de l'efficience des ressources investies; ou démontrer l'apport des CI à répondre aux politiques sociales, nationales et/ou institutionnelles; ou bien, démontrer la qualité des démarches entreprises et favoriser l'imputabilité des acteurs. Dans notre cas, il s'agit du type de preuves permettant d'atteindre l'objectif général de cette recherche. Laveault et Grégoire précisent que l'instrument ou les données ne peuvent être valides en général (2002). Au contraire, seules sont valides les inférences en faveur desquelles suffisamment d'arguments et de données empiriques ont pu être rassemblés pour les soutenir

(Messik, 1989). Aussi, il est retenu que la démonstration en faveur de la validité constitue une accumulation d'arguments ou de preuves.

2.1.1. Les types de validités

La validité est un concept qui a été opérationnalisé de multiples façons, en fonction des objectifs poursuivis. L'état des lieux sur ce sujet proposé par André, Loye et Laurencelle (2015) fait ressortir le caractère polysémique du concept de validité. Sans contester d'autres conceptualisations, cette étude positionne simplement la validité du QuizCI dans son contexte spécifique, pour les usages souhaités. Nous avons à cet effet retenu la typologie de la validité proposée par Laveault et Grégoire(2002)⁵ afin d'organiser l'information sur la validité. Elle permet de catégoriser les différents types d'arguments et informations destinés à documenter la validité. Aussi, elle permet de positionner la validité de l'instrument dans son contexte empirique, une réalité complexe qui implique différents points de vue.

Laveault et Grégoire identifient les types : (a) validité apparente, (b) validité de contenu, (c) validité en référence à un critère concomitant ou prédictif et (d) validité conceptuelle. Leur typologie est reproduite au tableau 3.

Tableau 3

Types de validité

| Type | Description |
|--|---|
| Apparente | Évaluation superficielle de la qualité des items |
| Contenu | Évaluation formalisée de la qualité des items par des experts |
| En référence à un critère : concomitante | Évaluation du degré de corrélation des scores à l'item ou à l'échelle avec une mesure prise comme référence |
| En référence à un critère : prédictive | Évaluation de la prédiction d'une observation future réalisée à partir des scores à l'item ou à l'échelle |
| Conceptuelle ou théorique | Évaluation du sens à attribuer aux scores à l'item ou à l'échelle sur la base d'un modèle théorique |

Source : Laveault et Grégoire, 2002, p. 165

⁵ La version 2002 est bien adaptée à notre propos. Aussi, l'usage de la version 2015 apparaît non requis.

Chaque type (apparente, de contenu, en référence à un critère et conceptuelle) regroupe les preuves de même nature. Ainsi, ce cadre permet de systématiser l'organisation de la preuve de validité en fonction de son apport et de mettre en perspective le rôle des différents types de validité au cours du processus de validation. Les preuves peuvent provenir d'arguments théoriques, dégagés de constats logiques; ou encore, elles peuvent être empiriques, soit produites à partir des données collectées avec instrument en question (Laveault et Grégoire, 2002). Ainsi, la typologie de Laveault et Grégoire met en perspective l'apport complémentaire des types de preuves pouvant être apportées tout au long d'une démarche de validation. Plus spécifiquement, elle permet de comprendre la différence ou l'écart dans le caractère convainquant de l'information proposée en regard de la démonstration de validité dans une variété d'études.

2.1.1.1. La validité apparente

Les arguments en faveur de la validité apparente sont recherchés du côté de la population ciblée par la mesure, par exemple les étudiants qui ont répondu au QuizCI. Les dispositifs retenus sondent l'accueil réservé à l'instrument et la compréhension des formulations ou de la terminologie utilisée. Cette information peut permettre d'améliorer l'instrument ou encore de documenter son acceptation sociale (Laveault et Grégoire, 2002). Les auteurs précisent que l'information obtenue avec ce type de dispositifs doit être mise en contexte avec l'expérience de l'étudiant.

2.1.1.2. La validité de contenu

Les dispositifs pour produire des arguments en faveur de la validité de contenu font appel à des experts pour savoir si le questionnaire leur paraît valide. L'expert est choisi en fonction de son expertise. Il est souvent invité lors de l'élaboration d'un instrument à réagir à son contenu ou à l'efficacité du format de l'instrument proposé. Il pourra identifier des problèmes dans la formulation des items, dans la structure de l'instrument, suggérer des améliorations quant à la couverture de l'objet mesuré, ou sanctionner l'instrument en regard de son expertise. Cette information contribue à améliorer l'instrument et à obtenir l'assentiment des acteurs intéressés: la reconnaissance des pairs, l'approbation des autorités compétentes ou celle d'interlocuteurs spécialisés (Laveault et Grégoire, 2002). Les auteurs rappellent que l'information colligée avec de tels dispositifs peut être teintée du biais de l'expert consulté.

2.1.1.3. La validité en référence à un critère

La preuve de validité en référence à un critère repose sur l'hypothèse que si un instrument mesure un trait précis et défini, le résultat (score au questionnaire par exemple) devrait être suffisamment corrélé avec le résultat obtenu d'un autre instrument mesurant le même trait ou un trait voisin (Laveault et Grégoire, 2002).

La preuve de validité en référence à un critère concomitant consiste à mettre en perspective des résultats comparables pour un même répondant. Il s'agit par exemple de comparer pour un même groupe l'ordre des scores totaux à deux épreuves censées mesurer la CI. Si le score de l'étudiant occupe sensiblement le même rang dans son groupe dans les deux cas, par exemple le même quartile, cette correspondance constitue un argument en faveur du fait que les deux instruments mesurent le même objet. La preuve de validité en référence à un critère externe peut être illustrée par une droite de régression linéaire. Selon le mode de représentation, si les données des deux instruments présentent des droites identiques ou parallèles (donc avec une pente similaire) ou si les scores pour un même étudiant présentent une intercorrélation suffisante, la concomitance des données peut être établie (Laveault & Grégoire, 2002a, p. 196).

La preuve de validité en référence à un critère peut provenir d'une corrélation entre deux mesures différées dans le temps. L'instrument ainsi validé peut agir alors comme prédicteur. Ainsi, un score permet de prédire un résultat ultérieur. Laveault et Grégoire présentent l'exemple de « prédire la taille à l'âge adulte (critère), par la mesure de la taille de l'enfant à quatre ans (prédicteur) » (Laveault & Grégoire, 2002a, p. 178). Suite à la mesure de ces deux variables auprès d'un grand échantillon, une corrélation significative entre les données constitue un argument en faveur de la validité prédictive des tendances dans la relation perçue entre ces variables. Cette démonstration s'appuie sur la stabilité et la généralisabilité des données (Scallan, 1981).

Cependant, ce type de preuve se constitue avec rigueur et vigilance. Lors de l'usage de ce type de dispositifs, la littérature émet trois mises en garde au chercheur néophyte que nous sommes. D'abord, Laveault et Grégoire (2002a) notent que les inférences faites à partir de la preuve de validité en référence à un critère impliquent des considérations morales, sociales et éthiques, entre autres pour les conséquences du testing sur le sujet. En bref, il faut garder à

l'esprit qu'il s'agit d'une généralisation. Par exemple, il serait erroné d'affirmer qu'un enfant petit à quatre ans sera petit adulte. Il a plutôt une certaine probabilité de l'être.

Ensuite, Penta, Arnoud et Decruynaere (2005) soulignent que pour convaincre du caractère stable et généralisable des données, cette preuve requiert un très grand échantillon de données, récoltées en plusieurs temps et auprès d'un grand nombre de sujets provenant de la population visée. Il s'agit d'un résultat pouvant être documenté, par exemple, par une étude longitudinale. Ainsi, ce type de preuve est plus rarement constitué, en raison des ressources et du temps requis pour la colliger.

Finalement, Borsboom, Mellenbergh et van Heerden (2004) ajoutent que les inférences faites à partir d'une table de coefficients de corrélations entre le score total et un autre score reste une preuve insuffisante. Il s'agit d'une preuve circonstancielle et dépendante de l'échantillon en question. Pour que cette preuve soit valide de façon absolue, le chercheur doit avoir établi au préalable des preuves du caractère analogue sur le plan conceptuel des deux corpus de données, ainsi que des preuves du caractère comparable des données d'un point de vue métrique. Ces deux types de preuves relèvent de la validité conceptuelle.

2.1.1.4. La validité conceptuelle

La démarche pour documenter la validité conceptuelle d'un instrument vise à vérifier que le résultat à l'épreuve, dans le cas du QuizCI il s'agit du score au questionnaire, supporte le modèle théorique sur lequel s'est appuyée son élaboration. Le QuizCI a été élaboré dans le but de mesurer la CI. Aussi, les items qui le composent représentent une conceptualisation de la CI et les données que l'instrument collecte doivent la représenter.

La preuve de validité de type conceptuel décrite par Laveault et Grégoire (2002) peut être documentée à l'aide de multiples dispositifs. Ces dispositifs constituent souvent des analyses quantitatives faites à partir des données collectées avec l'instrument. Ces analyses produisent de l'information à partir de données empiriques. Ces informations documentent les caractéristiques et les propriétés des données. Puis, ces informations peuvent permettre de faire des inférences sur la validité de l'instrument ou sur l'usage des données.

En fait, les auteurs Borsboom, Mellenbergh et van Heerden (2004) présentent des arguments à l'effet que la validité conceptuelle est la seule vraie preuve de validité. Pour l'expliquer, les auteurs (Borsboom et al., 2004)(Borsboom et al., 2004)(Borsboom et al., 2004)(Borsboom et al., 2004)(Borsboom et al., 2004)(Borsboom et al., 2004)circonscrivent les enjeux de la validité autour de deux conditions : un instrument est valide pour la mesure d'un attribut si, premièrement, le trait existe et si, deuxièmement, les variations du trait causent les variations du résultat de la mesure. Cette expression simple et claire du concept de validité n'assouplit en rien les exigences des démarches pour en documenter la preuve.

Pour la documenter, les trois auteurs énoncent eux aussi que la preuve de la validité [conceptuelle] se collige par une accumulation d'information en un même sens. Ils ajoutent cependant que la preuve de validité [conceptuelle] doit permettre de répondre à des questions relevant de trois domaines. Est-ce que le QuizCI réagit en fonction de la mesure captée? Est-ce que le QuizCI représente ce qu'il est censé mesurer? Est-ce que les données collectées avec le QuizCI mesurent la CI?

2.1.1.4.1. Est-ce que le QuizCI réagit en fonction de la mesure captée?

Premièrement, le chercheur peut démontrer l'existence de la compétence mesurée qui s'exprime lorsqu'un stimulus (les items) la cause. La preuve de validité doit montrer que le test valide permet de n'exprimer rien de moins, ni rien de plus que le trait mesuré; par exemple la CI. Ainsi, les preuves en faveur de la validité visent à mettre en évidence l'existence du trait mesuré dans une position dite ontologique (Borsboom et al., 2004). Une définition générique de l'étude ontologique résume qu'il s'agit de colliger un ensemble d'informations définissant le concept et permettant de décrire ses principales composantes et leur organisation interne (Centre national de ressources textuelles et lexicales, 2015). À cet effet, André, Loye et Laurencelle (2015) rappellent que la théorie classique des tests propose l'étude de la cohérence interne du contenu du test, celle-ci généralement étudiée à l'aide du coefficient alpha de Cronbach (1951). Lorsque la cohérence interne est suffisante, l'erreur de mesure a si peu d'impact sur le score total que l'instrument est réputé permettre de collecter d'une fois à l'autre, auprès d'échantillons similaires, des données comparables.

2.1.1.4.2. Est-ce que le QuizCI représente la CI dans sa nature et sa structure conceptuelle?

Deuxièmement, le chercheur peut tenter de fournir des preuves de la représentation conceptuelle à laquelle réfère la variable mesurée dans sa nature et sa structure conceptuelle (Borsboom et al., 2004). Il s'agit d'examiner la dimensionnalité de l'objet capté par l'instrument, en recherchant certaines correspondances entre la conceptualisation préconisée théoriquement et les données empiriques collectées avec celui-ci. Les analyses factorielles ou structurelles sont dans ce cas-ci appropriées (André et al., 2015).

2.1.1.4.3. Est-ce que les données collectées avec le QuizCI mesurent la CI?

Troisièmement, le chercheur peut chercher des preuves afin de comprendre quel lien entretient le trait avec sa mesure et pouvoir inférer ou documenter ses effets et ses influences (Borsboom et al., 2004). Pour ce faire, le chercheur adhère à l'hypothèse que le trait mesuré existe de façon latente chez tous les sujets de la population visée. L'item offre une opportunité d'exprimer le trait latent et la réponse à l'item constitue l'expression de ce trait. L'instrument permet de matérialiser le trait mesuré et de capter ses variations, chez les sujets de l'échantillon. La preuve en faveur de la validité repose sur la démonstration que ce trait peut effectivement être mesuré par cet instrument en particulier, car le score au test varie effectivement au gré des variations du trait latent. Pour produire des preuves de ce domaine de la validité, les méthodes issues de la théorie de réponse à l'item sont ici préconisées (André et al., 2015).

L'information fournie par les différents dispositifs mentionnés pour documenter les types de validité a été éprouvée dans différentes expériences de mesure de la CI. Cette littérature permet de connaître l'usage qui en est fait et leur apport respectif pour documenter la validité.

2.2. Recension des écrits étudiant la validité

Une recension des écrits permet de connaître les dispositifs utilisés par les auteurs pour produire l'information visant à documenter la validité du QuizCI. Ont été retenues les études décrivant un instrument de collecte de données sur la CI d'étudiants post-secondaire et proposant au moins un dispositif pour documenter la validité de ses données. Cette recension a débuté à travers la littérature québécoise. Onze documents ont répondu à ces critères. Les dispositifs recensés permettent de documenter la validité de type apparente et de contenu. Aussi,

nous avons poursuivi la recension des écrits dans la littérature non québécoise traitant d'un instrument collectant des données sur la CI d'étudiants universitaires et proposant au moins un dispositif pour documenter la validité conceptuelle. Dans la littérature anglophone, au-delà d'une centaine d'études ont été repérées (Scopus, 10 juin 2013). Parmi celles-ci, cinq études ont été retenues, car elles documentent les trois domaines de la validité décrits.

2.2.1. Les expérimentations québécoises

Le corpus québécois est composé de onze documents qui décrivent l'usage d'un instrument de collecte de données sur la CI des étudiants et proposent au moins un dispositif pour montrer la qualité des données que chacun permet de collecter. Ce sont dix textes (Allard et Gérin-Lajoie, 2011; April et Beaudoin, 2006; Bourgeois, Pouliot et Ruest, 2011; Du Ruisseau, Gamache-Vaillancourt, Mallet, Roy et Hiller, 2010; Durnin et Fortier, 2008; Gagnon-Mountzouris, 2011; Gervais et Arseneault, 2005; Lebrun et al., 2007; Mittermeyer, 2005; Simard, 2009) et un document audiovisuel présentant une communication orale dans un congrès (Gagnon-Mountzouris, 2011). Le tableau 4 présente les onze documents. Trois constats généraux s'en dégagent.

Premièrement, chaque expérimentation se déroule dans un contexte spécifique et unique. Les caractéristiques de l'échantillon sont systématiquement dissemblables : tous issus de disciplines et de cycles d'études variés; comptant d'une trentaine (Gervais et Arseneault, 2005) à plus de 3 000 sujets (Mittermeyer et Quirion, 2003). Dans ce même sens, l'instrument utilisé pour collecter des données dans chaque cas est adapté au contexte et à l'environnement informationnel. Aussi, le caractère spécifique de l'expérimentation et de l'instrument semble difficilement contournable en raison du contexte universitaire.

Deuxièmement, l'ensemble des instruments jouit d'une reconnaissance locale. Chacun connaît une utilisation périodique et répétée, dans un contexte formel.

La première étude présentée au tableau 4 est l'enquête subventionnée par la CRÉPUQ et menée par deux professeures-chercheuses de l'Université McGill, Diane Mittermeyer et Diane Quirion (2003). Cet instrument a été utilisé une seule fois au Québec pour capter un portrait général de la CI des étudiants entrant à l'université. Il a toutefois été repris en Belgique (Thirion

et Pochet, 2009), au Royaume-Uni (Lonsdale et Armstrong, 2006), en Australie (Hughes, 2010) et à l'Université de Madrid (Arias Coello et Simón Martín, 2008). De plus, cette étude est largement citée au Québec et dans la francophonie (Lefevre et Grenier-Gire, 2011). Elle a été reconnue et répétée.

Six des instruments contribuent à évaluer les étudiants dans le cadre d'un cours crédité (April et Beaudoin, 2006; Bourgeois et al., 2011; Durnin et Fortier, 2008; Gervais et Arseneault, 2005; Lebrun et al., 2007; Simard, 2009); alors que les quatre autres ont une fonction certificative dans un programme d'études universitaires (Allard et Gérin-Lajoie, 2011; Du Ruisseau et al., 2010; Dumont, Dupuis, Foucault, Hiller et Proulx, 2005; Gagnon-Mountzouris, 2011). L'utilisation périodique et répétée de chaque instrument dans un contexte formel constitue un argument en faveur de la crédibilité des instruments. Le troisième constat peut être à l'effet que tous les instruments collectent des données dichotomiques ou polytomiques ordinales, et non pas d'intervalle ou de ratio. A priori, ces données ne présentent pas les qualités métriques requises pour l'usage de certaines méthodes quantitatives utilisées pour documenter la preuve de validité, par exemple des analyses corrélationnelles contribuant à documenter la preuve en référence à un critère.

Tableau 4

Documents québécois recensés

| Études | Échantillon | Modalités de collecte | Nature des données |
|--|--|--|--|
| Mittermeyer & Quirion, 2003 | 3000 étudiants Toutes disciplines confondues Pré université | Épreuve [Questions de connaissance et d'habileté à choix multiples proposant des leurres et une bonne réponse] | Quantitatives dichotomiques |
| Durnin & Fortier, 2008 | 150 à 300 étudiants Disciplines 1er cycle dans les universités québécoises | Questionnaire d'auto-évaluation des connaissances composé de : 150 questions traduites ou adaptées de SAILS; 35 questions issues de Mittermeyer et Quirion (2003); 15 autres items | Qualitatives ⁶ et quantitatives dichotomiques |
| Bourgeois, Pouliot, & Ruest, 2011 | Environ 50 étudiants Toutes disciplines 2 ^e cycle | Questionnaire de connaissances et de compréhension | Quantitatives dichotomiques |
| Dumont, Dupuis, Foucault, Hiller, & Proulx, 2005 | 90 à 160 étudiants Disciplines Cycles supérieurs | Portfolio (Du Ruisseau, et al.) | Qualitatives |
| Du Ruisseau, Gamache-Vaillancourt, Mallet, Roy, & Hiller, 2010 | 1800 étudiants Génie, cycles sup. | Portfolio (Dumont, et al.) 2 questionnaires formatifs; 1 questionnaire sommatif de connaissances et de compréhension | Qualitatives et quantitatives dichotomiques |
| Gagnon-M., V., 2011 | Environ 50 étudiants; cours MTR8 (génie) Cycles supérieurs | Questionnaire de connaissances, de compréhension et de tâche complexe | Quantitatives dichotomiques et polytomiques |
| Lebrun, Perreault, & Verreault, 2007 | Environ 50 étudiants Éducation, 1 ^{er} cycle | Questionnaire d'auto-évaluation des apprentissages sur échelle d'attitudes | Quantitatives polytomiques |
| April & Beaudoin, 2006 | 78 étudiants Éducation, 1 ^{er} cycle | Observations; description qualitative | Qualitatives |
| Gervais 2004 | 24 étudiants Éducation, 1 ^{er} cycle | Questionnaire sur les habitudes de recherche; observation directe | Quantitatives polytomiques |
| Simard, 2009 | 1500 étudiants Génie, cycles sup. | Observations en laboratoire informatique | Qualitatives |
| Allard & Gérin-Lajoie, 2011 | 500 étudiants ; toutes disciplines; cours : 1037PREPA; | Questionnaires et tâche complexe à réaliser sur poste informatique | Quantitatives polytomiques |

⁶ Les données qualitatives peuvent être recodées et traitées comme des données quantitatives ordinales.

2.2.2. Les études complémentaires

Le second corpus est composé d'études présentant une collecte de données sur la CI d'étudiants universitaires ayant pour objectif général de documenter la validité de leur instrument et des données qu'il collecte. Comme il a été mentionné, dans la littérature anglophone, au-delà d'une centaine d'études scientifiques répondent à ces critères (Scopus, 10 juin 2013). Aussi, nous avons ciblé celles annonçant documenter la validité conceptuelle à l'aide entre autres de méthodes de la TRI. Ce dernier critère a réduit le corpus à cinq études présentées dans le tableau 5. Ce résumé permet de constater que le contexte général de collecte de données des études complémentaires s'avère comparable aux études québécoises.

Premièrement, leurs contextes sont variés. Les données ont été collectées auprès d'échantillon provenant de divers disciplines et cycles d'études, tandis que la taille de leur échantillon respectif s'étend de 350 participants (Timmers et Glas, 2010) à des milliers de sujets (DeMars et al., 2003; Katz, 2007b; Mery et al., 2011; O'Connor et al., 2001). Encore ici, leur environnement informationnel respectif est distinct et propre à chaque université.

Deuxièmement, chacun des instruments a connu un usage périodique et répété dans un contexte formel :

- L'instrument ICT (Timmers et Glas, 2010) est devenu en 2010 une évaluation certificative obligatoire au programme des *Sciences des communications* et à celui de *Marketing international* de l'University Twenty des Pays-Bas.
- L'instrument ISST constitue également une activité d'évaluation certificative depuis 2003, obligatoire dans tous les programmes de la *James Madison University* (DeMars et al., 2003).

Les trois autres instruments ont été intégrés au cursus de dizaines d'établissements universitaires américains :

- l'instrument iSkills Assessment (iSkills) pour le réseau des Arizona States Universities;
- l'instrument Online Research Lab (ORL) pour le réseau des California State Universities;
- alors que le projet Standardized Assessment of Information Literacy Skills (SAILS) a été utilisé dans 80 universités nord-américaines (Kent State University, 2015).

Troisièmement, les instruments présentés dans les études complémentaires collectent aussi des données dichotomiques ou polytomiques dichotomisées et non pas d'intervalle ou de rapport.

Tableau 5

Cinq études complémentaires recensées

| Études | Instrument | Échantillon | Modalités de collecte | Nature des données |
|-----------------------|---|--|---|-----------------------|
| Katz, 2007 | iSkills ETS's iSkills Assessment créé pour les California State Universities | Validé sur 1 000 étudiants à l'été 2004; en 2007, administré à 63 établissements post-secondaires de l'état de Californie. | Situation d'apprentissage en ligne; environnement numérique sophistiqué; étudiant doit performer des recherches pour répondre; Durée approximative de 75 minutes en 2 parties (35 + 45): 1ere partie, obligatoire lors de l'admission, soit un questionnaire d'auto-évaluation de leur compétence avant de passer les tests; 2e partie, évaluation sommative obligatoire en fin de 1ere année universitaire. Évaluation certificative | données polytomiques |
| Mery et Newby, 2010 | ORL Online Research Lab (ORL) créé pour l'Arizona State University | En 2009-2010 auprès de 1,400 étudiants de multiples disciplines | Cours en ligne crédité sur deux semestres; projet pilote dans le English Department; pré et post-test. 21 tâches; exigence institutionnelle; Évaluation certificative | données dichotomiques |
| Timmers et Glas, 2009 | ICT ICT Skills créé pour l' University Twenty (Netherlands); | En 2006-2007, 345 étudiants; soit 193 en Sciences des communications et 152 en Marketing international | Situation d'apprentissage intégrée à un cours obligatoire de 1ere année, 1ere session du baccalauréat. Situation d'apprentissage en classe et en bibliothèque recourant à la recherche sur ordinateur. 13 rencontres; évaluation formative, suivie d'une évaluation sommative. Cours obligatoire de méthodologie. Évaluation certificative | données polytomiques |

(Tableau se continu)

| Études | Instrument | Échantillon | Modalités de collecte | Nature des données |
|--|--|--|--|-----------------------|
| DeMars, Cameron et Erwin, 2003 | ISST Information Seeking Skills test-ISST créé pour la James Madison University, (UK) | Cette université compte environ 15 000 étudiants. L'instrument a été validé auprès des étudiants <i>seniors</i> de la cohorte 1998-1999. Son usage est récurrent depuis 1999-2000. | 8 modules d'apprentissage en ligne: <i>Go for the Gold</i> . L'étudiant doit performer des recherches pour répondre de façon autonome, à livres ouverts. Le nombre de reprises illimitées; durée approximative de 40 à 60 minutes. Questionnaire objectif et tâches complexes axées sur la prise de décision; environnement numérique d'apprentissage (ENA) sophistiqué; Réussite obligatoire avant la fin de la 1 ^{re} année universitaire; sans quoi un cours doit être suivi. Évaluation certificative | données dichotomiques |
| O'Conno r, Radcliff, et Gedeon, 2001; 2002 | SAILS Standardized Assessment of Information Literacy Skills (SAILS) créé pour la Kent State University (US) | 537 étudiants en 2 ^e année; Échantillon homogène; Demande des échantillons de >500 | Questionnaire objectif en ligne; durée illimitée; (durée moyenne de 35 à 55 minutes) Réussite obligatoire avant le début de la 3 ^e année universitaire [soit avant l'entrée au baccalauréat]; sans quoi un cours doit être suivi : Évaluation certificative | données dichotomiques |

2.3. Recension des dispositifs pour colliger la preuve de validité

Parmi les 16 textes retenus, nous avons relevé les dispositifs pour documenter la validité. Ces dispositifs sont regroupés par type de validité afin d'apprécier l'apport des différents types d'arguments ou de preuves. Parmi les 16 textes recensés, onze textes décrivent un dispositif pour récolter des arguments de validité apparente. Tous les textes présentent un dispositif visant à documenter la validité de contenu. Seulement deux études documentent la validité en référence à un critère. Puis, 5 études produisent des informations destinées à documenter la preuve de validité conceptuelle. La prochaine section passe en revue les dispositifs et met en perspective leur résultat en regard de la validité.

2.3.1. Dispositifs pour constituer la preuve de validité apparente

Onze textes décrivent l'usage d'un dispositif permettant la collecte d'arguments en faveur de la validité apparente. Pour ce type d'argument, contrairement aux méthodes quantitatives, il

n'y a pas de règle de décision. Aussi, nous avons retenu deux caractéristiques qui influent sur l'information colligée en vue de documenter la preuve de validité apparente, les caractères spécifique et direct. Le caractère spécifique du dispositif intervient lorsque l'auteur pose au répondant des questions portant spécifiquement sur des thèmes d'intérêt pour la validité apparente. Le caractère direct survient lorsque le chercheur interagit directement avec le répondant à propos de la validité. Ces caractéristiques peuvent influencer la qualité et la quantité d'informations récoltées. La figure 1 positionne les textes recensés selon ces deux axes. La première case en haut à gauche réfère aux textes présentant un dispositif générique et indirect; en bas à gauche, un dispositif indirect spécifique; en bas à droite, direct et spécifique; puis en haut à droite, direct et générique. Les dispositifs et leur apport sont décrits suite à la figure 1.

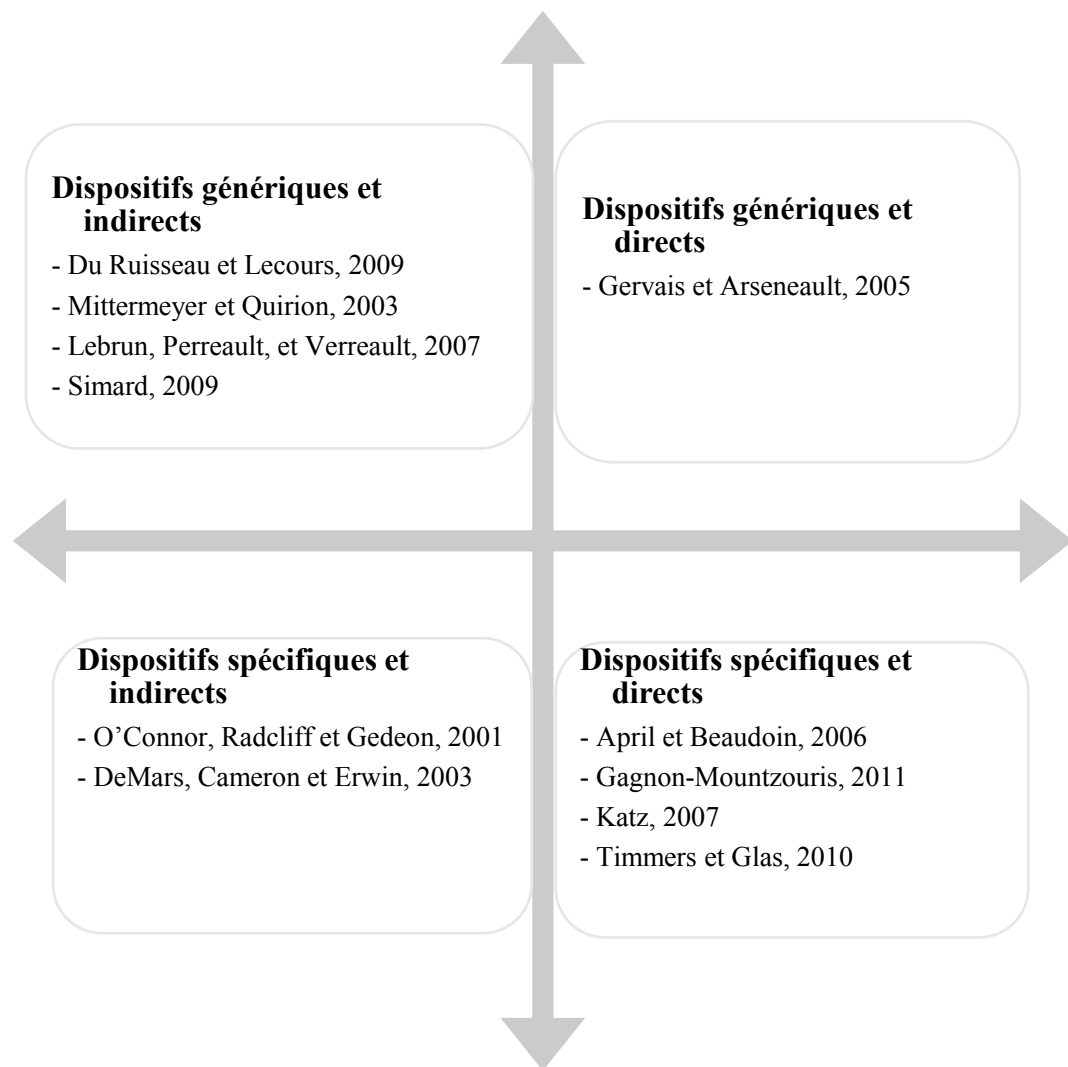


Figure 1 - Facteurs influençant l'étude de validation apparente.

Le dispositif générique et indirect a été mis de l'avant dans quatre des textes recensés. Il se résume à un espace pour les « commentaires », soit à même l'instrument (Du Ruisseau & Lecours, 2009; Mittermeyer & Quirion, 2003), ou en fin de questionnaire pour l'évaluation de l'enseignement par les étudiants (Lebrun et al., 2007; Simard, 2009). Ce dispositif ne vise pas spécifiquement à récolter des commentaires d'intérêt pour la preuve de validité de l'instrument, mais permet de le faire. Dans ces études, le chercheur n'interagit pas avec le répondant. Ce type de dispositif permet de récolter des informations inattendues, que le chercheur n'avait pas ciblées.

Le dispositif spécifique et indirect a été utilisé dans deux des études recensées :

- O'Connor, Radcliff et Gedeon (2001) proposent des items visant spécifiquement à documenter la validité apparente à même le test SAILS.
- Alors que DeMars, Cameron et Erwin (2003) profitent d'un sondage sur le service de bibliothèque pour adresser des questions portant spécifiquement sur ISST.

Ces items portent spécifiquement sur des thèmes d'intérêt pour l'étude de validation, ce qui favorise la collecte d'éléments de preuves pertinents pour documenter la validité apparente. Cependant, l'équipe de recherche n'échange pas directement avec le répondant à propos de la validité de l'instrument.

Un dispositif spécifique et direct a été utilisé dans quatre des textes :

- April & Beaudoin (2006) mènent une entrevue semi-structurée auprès de leur groupe-échantillon dans le cadre de la formation et recueillent leur perception à propos de l'épreuve.
- L'équipe de l'École des technologies supérieures documente la validité apparente de leur instrument lors d'une discussion avec quatre étudiants à la maîtrise ayant suivi la FD, soit le cours MTR801, de même que trois étudiants à la maîtrise n'ayant pas suivi la FD (Gagnon-Mountzouris, 2011).
- Katz (2007b), pour l'instrument iSkills, a dirigé un focus group auprès de trente-et-un étudiants de 1er cycle, soit l'échantillon visé.
- Timmers et Glas (2010) ont effectué une entrevue dirigée auprès de trois étudiants entrant au baccalauréat (Timmers & Glas, 2010), encore ici l'échantillon visé.

Le caractère spécifique de ces quatre dispositifs favorise la récolte de preuves sur des thèmes d'intérêt pour la validité. De surcroît, l'échange direct entre le chercheur et le répondant cible et explique l'objectif de l'exercice. De ceux-ci, trois notent que le contact direct avec les répondants leur a permis d'expliquer de façon plus explicite les objectifs et retombées attendues de la démarche de validation apparente (Gagnon-Mountzouris, 2011; Katz, 2007b; Timmers & Glas, 2010). Cette sensibilisation a pu faire varier en quantité et en qualité la collecte de l'information sur la validité de l'instrument.

Quatrièmement, le dispositif générique et direct a été utilisé par Gervais et Arseneault (2005). Ils ont noté les commentaires des étudiants durant une séance d'observation, alors que les étudiants procédaient à la résolution de problématiques informationnelles. Étant en classe, ils interagissaient directement avec les répondants. Les auteurs précisent ne pas être intervenus pour solliciter ou orienter les échanges.

Suite à la collecte d'arguments en faveur de la preuve de validité apparente, certains auteurs mentionnent comment ils ont traité l'information récoltée dans le but de l'interpréter en regard de la validité. Cinq textes présentent une analyse qualitative de la preuve recueillie (April & Beaudoin, 2006; Gagnon-Mountzouris, 2011; Gervais & Arseneault, 2005; Katz, 2007b; Timmers & Glas, 2010). Deux d'entre eux sont plus précis quant à la description du traitement de la preuve. April et Beaudoin (2006), présentent une analyse qualitative complète. Alors que Katz (2007b) décrit la méthode de l'Evidence-Centered-Design (EDC), qui a pour but de mettre en perspective les preuves sur la qualité des données. Cette méthode consiste en une triangulation des évidences sur la validité (apparente, d'experts et conceptuelle) pour établir un consensus sur la validité de l'instrument et des données qui en proviennent (Katz, 2004; Mislevy & Haertel, 2006).

2.3.2. Dispositifs pour constituer la preuve de validité de contenu

Tous les textes recensés présentent un dispositif pour colliger des arguments en faveur de la validité de contenu. Les auteurs de sept instruments rapportent avoir eu un échange verbal avec un ou d'autres bibliothécaires formateurs (April & Beaudoin, 2006; Bourgeois et al., 2009; Dumont et al., 2005; Durnin & Fortier, 2008; Gervais & Arseneault, 2005; Lebrun et al., 2007; Simard, 2009).

D'autres auteurs décrivent avoir mis en place un dispositif plus formel:

- Gagnon-Mountzouris (2011) relate des « rencontres préparatoires avec les professeurs; des évaluations à l'interne [auprès du personnel de la bibliothèque] et à l'externe auprès des membres d'un groupe spécialisé », le groupe de travail pour la Promotion du développement des compétences informationnelles.
- Timmers et Glas. (2010) procèdent à l'étude de validation d'experts de l'instrument ICT auprès d'un bibliothécaire, de deux étudiants de 3e cycle et de deux professeurs. Ceux-ci ont lu et commenté l'épreuve évaluative par courrier électronique.
- Katz (2007) a effectué l'étude de validation d'experts des données collectées avec l'instrument iSkills par focus group auprès d'étudiants de cycles supérieurs, de membres de l'équipe de bibliothèque et de membres du corps professoral.
- O'Connor, Radcliff et Gedeon (2001) ont sollicité six étudiants jugés expérimentés en CI pour l'étude de la validité d'expert. Ils ont demandé à leurs experts de distribuer les items en trois sous-groupes : les items faciles, de difficulté moyenne et les items difficiles. Les auteurs, suite à la collecte de données, comparent la difficulté attendue et celle observée. Les items dont les niveaux de difficulté attendu et observé présentaient un écart ont été révisés.
- Katz (2007b), Simard (2009), puis Timmers et Glas (2010) utilisent un tableau de spécification. Dans les trois cas, les auteurs de l'instrument ont classifié les items parmi les six niveaux cognitifs de connaissance énoncés par la Taxonomie de Bloom révisée (Anderson, 1992).
- Du Ruisseau et Lecours, (2009), à l'instar de Bourgeois, Pouliot et Ruest (2011) mentionnent avoir sollicité un expert en pédagogie universitaire. Du Ruisseau et Lecours (2009) précisent avoir consulté un bureau spécialisé en pédagogie pour l'« Instauration des rubriques (grille avec critères) pour l'évaluation avec la collaboration du Bureau d'appui pédagogique »;

En outre, six des instruments recensés ont été élaborés en collaboration avec un spécialiste en technologie des communications (Allard & Gérin-Lajoie, 2011; DeMars et al., 2003; Katz, 2007b; Lebrun et al., 2007; Mery et al., 2011; L. O'Connor et al., 2002). Les auteurs ont voulu

éviter qu'un aspect technique nuise à la validité des données. Mery et Newby (2011) racontent qu'ils ont sollicité un expert dans le but d'éviter que l'affichage trop lent ou incomplet des pages électroniques supportant leur instrument nuise à la compréhension de la tâche à effectuer.

La recension des dispositifs a permis de connaître la nature des expertises sollicitées pour documenter la validité d'un instrument de mesure de la CI.

2.3.3. Dispositifs pour constituer la preuve de validité en référence à un critère

Deux études recensées présentent une démarche visant à documenter une preuve de validité en référence à un critère. Il peut être rappelé qu'il s'agit de comparer des corpus de données censés prendre la même mesure en des temps différents ou des mesures sur deux instruments censés mesurer la même chose.

Mery et Newby en 2011 comparent, pour un même échantillon d'étudiants, le score à 20 items de l'instrument ORL avec le score à 20 items analogues du test SAILS⁷. Le résultat obtenu n'a pas permis de fournir de preuve de validité. Pour chaque couple d'items, les auteurs ne pouvaient pas les considérer suffisamment corrélés pour constituer une preuve que les deux items prennent la même mesure. Aussi, ce dispositif n'a pas permis de colliger de preuve de validité pour l'instrument ORL en référence à un critère par concomitance avec les données obtenues avec le test SAILS. Cette expérimentation illustre les doutes émis par Borsboom, Mellenbergh et van Heerden (2004) quant à la qualité de la preuve produite à l'aide de corrélations.

De leur côté, O'Connor, Radcliff et Gedeon (2001) mentionnent vérifier la concomitance entre le score total au test SAILS et la moyenne cumulative de l'étudiant à la fin de la 1^{ère} année universitaire, le *Grade Point Average (GPA)*. Elles ont vérifié si les scores pour un même étudiant se situent dans le même quartile. Les auteurs précisent avoir réalisé cette démarche pour

⁷ Rappelons que la validité du test SAILS a été documentée en 2001.

documenter la validité en référence à un critère concomitant, mais ne présentent pas l'information obtenue.

Enfin, nous remarquons que l'équipe du SAILS project (O'Connor et al., 2001), quinze ans après la publication de leur démarche, ont mis en place les conditions pour documenter une preuve de validité en référence à un critère prédictif. En effet, le test SAILS est largement administré. En 2005, il l'a été auprès de 80 universités nord-américaines, dont six établissements canadiens : York University à Toronto, University of British Columbia – UBC, University of Alberta, Brandon University au Manitoba, Western University à London, Ontario et l'University of New Brunswick (Adam et Maimets, 2005). De plus, la Kent State University, établissement instigateur, collecte des données annuellement depuis 2000, donc de façon longitudinale. Aussi, le *SAILS project* dispose d'un très grand échantillon de données récoltées en plusieurs temps dans un échantillon représentatif de la population visée. Puis, ils ont documenté le caractère métrique de leurs données en vue de produire de l'information sur la validité de leur instrument (Salem, 2005). Il est à noter que *documenter le caractère métrique de l'instrument* peut être effectué en documentant le troisième domaine de la validité conceptuelle présenté par Borsboom, Mellenbergh, et van Heerden (2004) et l'usage des méthode de la TRI.

2.3.4. Dispositifs pour constituer la preuve de validité conceptuelle

Cinq études recensées présentent des dispositifs pour documenter la validité conceptuelle. Ces dispositifs reposent principalement sur des méthodologies quantitatives. Les paragraphes suivants présentent le contexte de mesure et leurs démarches pour documenter la validité.

2.3.4.1. iSkills Assessment (Katz, 2007)

L'instrument iSkills Assessment est un questionnaire électronique se déroulant en deux parties, comportant respectivement 35 et 45 mises en situation. Pour chacune, l'étudiant choisit la réponse appropriée parmi 4 catégories. Les données ont été collectées auprès de 1000 étudiants à travers les California State Universities (Katz, 2007).

Pour documenter la preuve de validité conceptuelle, Katz présente l'estimé de la cohérence interne de l'instrument, le coefficient alpha de Cronbach. En fait, celui-ci cible les erreurs de mesure nuisant à la cohérence interne de l'instrument à l'aide d'une technique

d'estimation de l'homogénéité des items d'une mesure. Elle se base sur la corrélation des items entre eux, ainsi qu'entre eux et le score total (Laveault et Grégoire, 2002). Plus l'homogénéité des patrons de réponses est grande, plus l'échelle est fidèle au sens de la cohérence interne des items. Ce coefficient s'interprète selon la règle de décision proposée par Cronbach en 1951 :

| | |
|-------------------------------|----------------|
| α = inférieur à - 0,4] | insatisfaisant |
| α = [0,4 - 0,6] | faible |
| α = [0,6 – 0,8] | satisfaisant |
| α = supérieur à 0,8 | excellent |

Katz utilise le logiciel SPSS (IBM, 2010) pour calculer le coefficient de Cronbach. Au terme des ajustements, il en vient à un coefficient de fidélité $\alpha = 0,88$ pour l'instrument iSkills. Cet *excellent* coefficient alpha de Cronbach fournit une preuve en faveur de la validité conceptuelle de l'instrument iSkills. L'auteur explique qu'il s'agit d'une double preuve. D'abord, la valeur du coefficient constitue une preuve en faveur de la cohérence interne de l'instrument. Puis, le fait qu'il corrobore les arguments de validité apparente et de contenu présenté constitue une correspondance entre le résultat attendu et le résultat observé, soit un argument supplémentaire en faveur de la validité de l'instrument.

2.3.4.2. Online Research Lab -ORL (Mery, & Newby, 2010)

L'instrument ORL est un questionnaire électronique comportant 21 tâches. Pour chacune, l'étudiant doit répondre à des questions et choisir la réponse appropriée parmi 5 catégories. L'échantillon de données a été collecté auprès de 4500 étudiants à travers l'Arizona State University (Mery et al., 2011).

Mery, Newby et Peng (2011) documentent la preuve de validité conceptuelle de leur instrument par l'examen de 3 paramètres. Ils examinent si la difficulté de l'item est adéquate pour l'échantillon visé. Ils vérifient si les items permettent de discriminer les étudiants composant l'échantillon en fonction de leur CI. Puis, ils présentent le coefficient alpha de Cronbach pour estimer la cohérence interne de l'instrument.

Pour mettre en lumière la difficulté respective de chaque item, les auteurs présentent l'indice de simplicité (**IS**). Celui-ci se calcule tel que $IS = R/T$, où R= le nombre de bonnes réponses à cet item et T, nombre total de réponses. La difficulté de l'item est ainsi représentée

par un pourcentage correspondant à la proportion des sujets qui ont réussi l’item. Les auteurs ont ensuite classifié les items selon la valeur de l’indice, afin d’émettre des recommandations pour une révision de l’épreuve. Ils ont utilisé la règle de décision « Niveaux des items et recommandations pour la révision de l’instrument » d’Escouadro (2000) présentée au tableau 6. Cette règle comporte 8 niveaux de difficulté afin de déterminer si le niveau de difficulté de l’item est adéquat pour la mesure.

Tableau 6

Règle de décision de l’indice de simplicité (IS).

| Indice de simplicité | Niveaux | Interprétation |
|----------------------|-----------------------|--|
| > 0,80 | Extrêmement facile | L’item doit être retiré définitivement |
| 0,70 – 0,80 | Facile | L’item doit être révisé |
| 0,60 - 0,69 | Moyennement facile | L’item est retenu |
| 0,40 - 0,59 | Moyen | L’item est retenu |
| 0,30 - 0,39 | Moyen - difficile | L’item est retenu |
| 0,20 - 0,29 | Difficile | L’item est retenu |
| 0,10 - 0,19 | Très difficile | L’item doit être révisé |
| < 0,10 | Extrêmement difficile | L’item doit être retiré définitivement |

Source: Mery, Y., Newby, J., & Peng, K. (2011). Assessing the reliability and validity of locally developed information literacy test items. *Reference services review*, 39(1), p.106.

Mery, Newby et Peng (2011) ont utilisé l’indice de discrimination (*d*) pour prendre connaissance de la capacité des items à discriminer les étudiants en fonction de leur CI. L’indice de discrimination répartit les scores totaux à l’épreuve par quartile et fait le rapport entre la moyenne des scores totaux du groupe ayant les résultats les plus élevés (Re) avec le groupe ayant les résultats les plus faibles (Rf). Cet indice s’étend de 0 à 0,39, selon la règle de décision utilisée par Mery, Newby et Peng (2011). Afin d’émettre des recommandations pour une révision de l’épreuve, les auteurs ont classifié les items selon leur capacité de discrimination (*D*), encore ici selon la règle de décision d’Escouadro (2000), reproduite au tableau 7.

Tableau 7

Règle de décision de l'indice de discrimination (D).

| Indice de discrimination | Interprétation |
|--------------------------|--|
| $D \geq 0,40$ | L'item discrimine plutôt bien |
| $0,30 \leq D \leq 0,39$ | L'item n'a pas vraiment besoin de révision |
| $0,20 \leq D \leq 0,29$ | L'item a besoin d'un peu de révision |
| $D \leq 0,19$ | L'item doit être révisé en profondeur |
| $D \leq 0,0$ | L'item doit être éliminé |

Source: Mery, Y., Newby, J., & Peng, K. (2011). Assessing the reliability and validity of locally developed information literacy test items. *Reference services review*, 39(1), p.108.

Mery, Newby et Peng (2011) vérifient ensuite la cohérence interne de l'instrument ORL à l'aide du coefficient alpha de Cronbach pour chaque item. La 1^{re} version du questionnaire avait un coefficient moyen $\alpha = 0,62$. Pour l'améliorer, les chercheurs ont reformulé 40 des 125 questions. Suite à quoi, le coefficient atteint 0,82, indiquant une bonne cohérence interne.

2.3.4.3. Information / Communication / Technologies Skills (Timmers et Glas, 2010)

Les données ont été collectées à la Twenty University (Pays-Bas). L'instrument ICT collecte des données qualitatives obtenues avec une grille d'observation qui sont recodées en données polytomiques, ainsi que des données polytomiques collectées avec une échelle d'attitudes de type Likert. Cette étude vise à colliger des preuves que les données traduisent la conceptualisation théorique ayant soutenu l'élaboration de l'instrument, soit le second aspect de la validité conceptuelle. Pour y parvenir, ils proposent quatre dispositifs.

Premièrement, Timmers et Glas réalisent une analyse factorielle exploratoire à rotation Varimax. La solution factorielle retenue est composée de trente-deux items ayant suffisamment saturé sur deux facteurs, dont 26 items à 0,30 *ou tout près*. Les auteurs recherchaient une solution factorielle à 4 facteurs afin de correspondre à leur cadre conceptuel à quatre dimensions. Au final, le premier facteur comprend 20 items saturant à 0,30 *ou tout près*. Un deuxième facteur est composé de 6 items. Timmers et Glas constatent que les items composant le 1^{er} facteur correspondent aux items des dimensions 1 et 2 du cadre conceptuel. Les items composant le 2^e

facteur correspondent aux items de la dimension 3 du cadre conceptuel. Puis, les items de la dimension 4 du cadre conceptuel ne sont pas retenus dans la solution factorielle.

Deuxièmement, Timmers et Glas modélisent les données avec un *General Partial Credit Model (GPCM)*, à l'aide du logiciel LM (Microsoft, 2012). La taille de l'échantillon est de 345 sujets pour 32 items. Cette taille est suffisante pour un modèle de Rasch (Bond et Fox, 2007). Par ailleurs, les auteurs ne précisent pas quelle est la taille requise dans le cas d'un GPCM. Les estimés du modèle soutiennent également les inférences à l'effet que les items représentent trois des dimensions théoriquement attendues. Alors que les items de la 4^e dimension ne s'ajustent pas au modèle.

Troisièmement, les auteurs vérifient le lien qu'entretiennent les facteurs à l'aide d'une matrice de corrélations (r) de Pearson. Or, les items représentant les 3 premiers facteurs sont corrélés. Tandis que les items du 4^e facteur sont faiblement intercorrélés et ne semblent pas constituer un facteur. Ce résultat est cohérent avec l'information obtenue précédemment et cumule une preuve supplémentaire en faveur de l'inadéquation des items du 4^e facteur.

Quatrièmement, l'analyse de la cohérence interne appuie encore ces résultats. Pour les trois premières dimensions, les items par dimension présentent une cohérence interne qui se classe de satisfaisante à excellente ($\alpha_1 = 0,68$; $\alpha_2 = 0,74$; $\alpha_3 = 0,81$). Tandis que le coefficient pour les items de la 4^e dimension s'avère insatisfaisant ($\alpha_4 = 0,41$).

Au terme de la preuve colligée, il ressort que les items censés représenter la quatrième dimension ne permettent pas de le faire adéquatement. À partir de ce résultat, les auteurs concluent que des items devront être révisés afin de mieux représenter la 4^e dimension du construit théorique.

2.3.4.4. Information Seeking Skills test-ISST (DeMars, Cameron et Erwin, 2003)

ISST est un questionnaire « papier-crayon » de 40 items à choix multiple (DeMars et al., 2003). Les données ont été collectées auprès des étudiants à travers la James Madison University, (UK). Les auteures ont documenté la preuve de validité conceptuelle de l'instrument et de leurs données avec 4 paramètres reposant sur la Théorie classique des tests (TCT). Elles calculent la cohérence interne, l'erreur de mesure, la moyenne des scores et la capacité de

discrimination de l'item en fonction de la CI. Puis, avec les méthodes de la Théorie de réponse à l'item (TRI), elles documentent l'ajustement des données aux modèles avec le Infit et Outfit. Ces coefficients mesurent l'adéquation des données au modèle après que les paramètres ont été estimés pour l'échantillon de données. DeMars, Cameron et Erwin comparent leur évolution sur cinq ans.

La méthodologie mise de l'avant par chacun des dispositifs est décrite, ainsi que l'information qu'il doit apporter. Cependant, la règle de décision utilisée et l'interprétation du résultat ne sont pas présentées. Aussi, le lecteur comprend théoriquement ce qui est fait, mais la démonstration de validité demeure hermétique.

2.3.4.5. Standardized Assessment of Information Literacy Skills (SAILS)

Les données ayant fait l'objet de l'étude de la validité conceptuelle du test SAILS ont été collectées en 2000 auprès de 537 étudiants à travers la Kent State University (US). Il s'agit d'un questionnaire électronique, à ce moment-là comptant 50 questions. Pour chacune, l'étudiant doit choisir la réponse appropriée parmi 2 à 5 catégories de réponses (O'Connor et al, 2002). Notons que cette étude s'inscrit dans une plus large démarche de validation nommée *SAILS project* (<http://www.sails-project.eu/portal>).

Pour documenter la preuve de validité, un modèle de Rasch a été appliqué avec le logiciel Winsteps (Linacre, 2002a). L'application du modèle proposé par Rasch estime la probabilité de réussite de chaque item et de chaque personne (1960). Ces estimés positionnent les scores des personnes et des items sur une même échelle en unité Logit. Cette unité mesure deux paramètres, soit son niveau de présence chez le sujet (pour une compétence il s'agit de l'habileté), ainsi que son niveau représenté par l'item (sa difficulté). Ce modèle assume que la capacité de discrimination (D) de chaque item est égale à 1.

Les auteures ont construit la preuve de validité selon la démarche d'ajustement en 6 questions suggérées par Stone et Wright dans leur ouvrage *Best test Design* (1979). Les trois premières questions ciblent l'ajustement des items au modèle, alors que les trois dernières concernent l'ajustement du score des personnes. Pour répondre à chacune des questions, les auteures utilisent l'information fournie par le logiciel Winsteps.

1. Est-ce que l'intensité croissante de CI requise pour répondre adéquatement au questionnaire est perceptible?

Pour le vérifier, O'Connor, Radcliff et Gedeon examinent la carte des personnes et des items (2001). Elles représentent sur une même échelle de mesure les items et les personnes. Les auteures peuvent vérifier si les items se regroupent dans un intervalle raisonnable de difficulté. Selon les recommandations de Linacre (2002b), l'intervalle acceptable s'échelonne de -3 et +3 sur l'échelle Logit. Les items présentant un niveau de difficulté au-delà de ces valeurs sont considérés être des données problématiques, car elles affectent l'ajustement de l'ensemble des autres données au modèle.

2. Est-ce que le positionnement des items le long de cette l'échelle a du sens?

O'Connor, Radcliff et Gedeon (2001) expliquent que les items qui devraient mesurer un haut niveau de compétence et ainsi être plus difficiles se retrouvent dans le haut de l'échelle, alors que les items présumés plus faciles se situent dans le bas de l'échelle. Pour y répondre, les auteures doivent comparer l'ordre du positionnement des items le long de l'échelle avec l'ordre prévu lors de l'élaboration de l'instrument. Rappelons que les auteures avaient demandé à des experts de classer les items selon leur difficulté. Aussi, pour répondre à cette question, les auteures comparent les résultats attendus et obtenus.

3. Est-ce que les items collectent des données s'ajustant au modèle de Rasch?

Pour répondre à cette question, les auteures analysent les coefficients Infit et Outfit. Les valeurs standardisées (zstd) recherchées devraient se situer entre -2 et +2 selon Linacre (2002).

4. Est-ce que les personnes sont convenablement séparées le long de l'échelle?

Pour y répondre, les auteures consultent l'indice de fiabilité et celui de séparation des items. La valeur recherchée pour l'indice de séparation est inférieure à 2,0. La valeur recherchée pour l'indice de fiabilité est supérieure à 0,80.

5. Le positionnement de chaque score le long de l'échelle a-t-il un sens?⁸

⁸ La cinquième question contribue davantage à documenter la validité en référence à un critère que la validité conceptuelle.

Pour répondre à cette question, les auteures analysent les mesure Infit et Outfit. Les valeurs standardisées (zstd) recherchées devraient se situer entre -2 et +2 selon Linacre (2002).

6. Est-ce que la mesure est valide pour toutes les personnes ?

Pour y répondre, les auteures consultent l'indice de fiabilité et celui de séparation des personnes. Les valeurs recherchées sont encore une fois : pour l'indice de séparation inférieure à 2,0; pour l'indice de fiabilité supérieure à 0,80 (Linacre, 2002a).

En somme, O'Connor, Radcliff et Gedeon (2001) décrivent les dispositifs, l'information que chacun apporte et la règle de décision. Cependant, elles ne présentent pas le résultat de ces analyses pour le test SAILS. Ainsi, elles décrivent comment faire la preuve, mais ne démontrent pas la validité de l'instrument SAILS. La démarche elle-même est toutefois claire et systématique.

En complément d'information, McMillen et Deitering (2007) ne décrivent pas la méthodologie derrière le test SAILS, mais décrivent leur expérience de l'usage de cet instrument. Elles éclairent les limites de l'usage du test SAILS. Ils rappellent les limites connues, soit la limitation des items objectifs à représenter une compétence, ainsi qu'à fournir des indications pour l'amélioration de la mesure, dans l'administration ou le contenu de l'instrument. Les auteures réalisent des focus groups pour répondre à cette dernière limite. Puis, ils relèvent également la limite à mettre en perspective les différentes cohortes, puisque la preuve de validité en référence à un critère concomitant est possible seulement si les qualités métriques des données sont démontrées : « At the outset, we didn't fully understand that the changes from year to year would limit our ability to compare cohort results to the extent that they did. » (McMillen et Deitering, 2007, p. 21) Également, les auteures ont eu à faire face manque de constance de l'instrument. L'instrument a été intégré dans 121 groupes-cours. Certains enseignants ont ajouté ou retiré des items à l'épreuve. Les données alors collectées proviennent de différents instruments, présentant une cohérence interne différente. Ces échantillons de données ne sont pas homogènes et ne peuvent être traités comme un seul ensemble.

Pourtant, malgré ces difficultés, les auteures demeurent convaincues de l'apport de données empiriques sur la CI des étudiants puisque, à terme, elles permettront : de mesurer la CI des étudiants, de représenter l'évolution de la compétence au cours du parcours universitaire, ainsi que de fournir des rétroactions sur l'efficacité des actions entreprises.

2.4. Buts de la présente étude

La question générale de la présente étude est : comment documenter la validité du QuizCI, instrument visant à mesurer la CI des étudiants à l'UQO Saint-Jérôme?

La recension des écrits a permis de prendre connaissance des objectifs, des pratiques, ainsi que de certains résultats de chercheurs contribuant à y répondre. Elle a permis de constater que différents types de validité peuvent être documentés et que chaque type de validité fournit une preuve de nature différente en regard de la validité.

2.4.1. Validité apparente et de contenu

L'information obtenue par les dispositifs de validité apparente a permis aux chercheurs des études recensées de mieux connaître la compréhension des étudiants par rapport aux items. Encore, elle a permis de mieux comprendre l'instrument répond aux attentes des étudiants en regard du contenu ou de la difficulté de l'épreuve. L'information obtenue par les dispositifs de validité de contenu a permis d'obtenir les rétroactions d'expert : premièrement, à propos de l'adéquation du contenu de l'épreuve; deuxièmement, en regard du caractère optimal de l'épreuve dans l'environnement numérique.

Ainsi, les dispositifs documentant la validité apparente et de contenu produisent de l'information pouvant aiguiller le chercheur sur la validité de l'instrument. Cette information peut alimenter sa réflexion sur les usages possibles de données. Or, Borsboom, Mellenbergh et van Heerden (2004) notent qu'une fois les données collectées auprès de la population visée, le chercheur doit composer avec l'échantillon en sa possession. Dans un cas où les données ne permettent pas de réaliser les analyses voulues, il faut collecter à nouveau des données.

Cette réflexion nous amène pour la présente étude à identifier avant la collecte de données les analyses pressenties, dans le but d'en identifier les postulats et les exigences. Cette réflexion

nous permet de mettre en place des conditions favorables à la réalisation des analyses pressenties. Ces démarches sont réalisées lors de l'élaboration du QuizCI et décrites en section 3.3.

2.4.2. Validité en référence à un critère

Les dispositifs destinés à documenter la validité en référence à un critère peuvent produire des arguments ou, lorsque la validité conceptuelle est établie, des preuves. Or, à la lumière des arguments de Borsboom, Mellenbergh et van Heerden (2004), présentant des arguments à l'effet que la validité conceptuelle est la seule vraie preuve de validité; et dans l'objectif de documenter la validité du QuizCI, la présente étude s'emploiera à produire des preuves empiriques de validité. Cependant, lorsque la validité conceptuelle sera documentée, il sera des plus intéressant de collationner la mesure en des temps différents ou d'examiner les liens qu'elle pourrait entretenir avec d'autres variables, par exemple à l'occasion d'études ultérieures.

2.4.3. Validité conceptuelle

Ainsi, à l'aide des méthodologies décrites dans les textes recensées, la présente étude procédera à l'étude de la validité conceptuelle du QuizCI afin de produire des preuves empiriques de sa validité, c'est-à-dire fondée sur des observations réelles, les données collectées.

Borsboom, Mellenbergh et van Heerden (2004) ont identifié trois domaines de la validité, soit premièrement, des preuves de nature ontologique; deuxièmement, des preuves de la structure dimensionnelle; troisièmement, des preuves que la CI en tant que trait latent, peut effectivement être mesuré adéquatement par le QuizCI. Pour documenter les trois domaines de la validité conceptuelle, les auteurs recensés ont mis en œuvre différentes méthodologies. Le premier domaine de la validité a été documenté à l'aide des méthodes de la Théorie classique des tests décrites par Katz (2007a) et Mery, Newby et Peng (2011). Le second domaine de la validité a été documenté par Timmers et Glas (2010). Enfin, le troisième domaine de la validité a été examiné par DeMars, Cameron et Erwin (2003), ainsi qu' O'Connor, Radcliff et Gedeon, (2001).

Les différents dispositifs recensés documentant les trois domaines de la validité produisent de l'information. Cette information empirique peut contribuer à répondre à la question générale de recherche. Aussi, les textes recensés nous fournissent des indications sur comment documenter la validité du QuizCI. Il s'agit de reproduire les analyses recensées, en les adaptant à notre contexte spécifique.

2.5. Questions spécifiques de recherche

Ainsi, pour vérifier l'apport des dispositifs et des méthodologies recensés et ainsi répondre avec justesse à la question : comment étudier la validité du QuizCI, la présente étude réalisera un processus de validation sur les données collectées avec le QuizCI. L'information produite par les analyses réalisées permette de dégager des éléments de preuves de validité ou de non-validité.

Pour ce faire, au cours de la phase préparatoire les dispositifs pour documenter la validité apparente et de contenu seront mis en place afin de favoriser la collecte d'un échantillon favorisant la réalisation des analyses destinées à documenter la validité conceptuelle. Puis, nous collecterons des données avec l'instrument ayant subi un processus de validation dans le but d'adresser les questions de recherche spécifiques (Qs) suivantes⁹:

Qs1 : Est-ce que le QuizCI réagit en fonction de la mesure captée?

Qs2 : Est-ce que le QuizCI représente la CI dans sa nature et sa structure conceptuelle?

Qs3 : Est-ce que les données collectées avec le QuizCI mesurent la CI?

Enfin, pour clore ce second chapitre remarquons qu'il incombe à l'auteur de démontrer comment cette information contribue à la preuve de validité. La littérature fait consensus sur le fait que la preuve de validité conceptuelle repose sur l'accumulation d'information/preuve dans un même sens, ainsi que sur la clarté de la nature de l'information obtenue. La variété ou l'exhaustivité de la nature des informations colligées, ainsi que la clarté de la règle de décision du dispositif peuvent y contribuer.

⁹ Pour ce faire, la section 3.6. de la présente étude décrit la méthodologie pressentie. Les analyses réalisées et leurs résultats sont présentés au quatrième chapitre, puis discutés au cinquième.

Chapitre 3 : cadre méthodologique

Ce troisième chapitre présente la phase préparatoire de l'expérimentation au cours de laquelle des dispositifs pour documenter la validité apparente et de contenu ont contribué à l'élaboration du QuizCI. Il décrit la collecte de données et l'échantillon destiné à l'expérimentation. Puis, il expose les analyses qui seront réalisées pour produire les informations desquelles pourront être dégagés des arguments ou des preuves de validité du QuizCI.

3.1. Type de recherche et justifications

Cette recherche se caractérise comme étant à la fois exploratoire et descriptive.

Elle est exploratoire particulièrement au niveau de l'application du concept de validité à un instrument mesurant la CI, ainsi que de par l'usage de méthodologies pour produire des arguments et des preuves de validités sur un instrument mesurant la CI, et ce, particulièrement en contexte universitaire canadien et francophone.

Par ailleurs, elle est descriptive de par le type de connaissances produites afin de : mettre en évidence l'existence de la CI dans une position dite ontologique; fournir des preuves de la représentation conceptuelle à laquelle réfère la variable mesurée dans sa nature et sa structure conceptuelle; ainsi que dans le but de démontrer que le score au QuizCI varie effectivement au gré des variations de la CI de l'étudiant.

3.2. Visées de la recherche

En ce sens, cette étude vise en premier lieu à documenter la validité du QuizCI et des données qu'il collecte pour l'usage qui en est fait dans le cadre des activités académiques de l'UQO. En second lieu, elle vise à proposer une démarche explicite, transparente et convaincante d'étude de validation. En troisième lieu, elle vise à présenter différentes méthodes disponibles et leurs apports aux Sciences de l'information.

3.3. Élaboration du QuizCI

L'élaboration du QuizCI s'est déroulée de 2009 à 2012. Par ailleurs, notre démarche a commencé à être instrumentée en 2011. Dix dispositifs ont été mis de l'avant pour préparer la

collecte de données viables pour les analyses pressenties. Ces dispositifs sont décrits dans les paragraphes suivants et résumés au tableau 9.

Les items du QuizCI ont été tirés de la traduction québécoise d'items du test SAILS, soit de la *Banque de questions : Connaissances informationnelles et habitudes de recherche* (Direction des bibliothèques de l'Université de Montréal, 2006). Plusieurs de ces items proviennent d'une traduction du test SAILS (Durnin et Fortier, 2008). L'idée de départ était de reproduire l'expérience du *SAILS Project* (Kent State University, 2015; Salem, 2005). Il s'agit d'instrumenter l'étude de la validité du QuizCI; d'une part, pour connaître sa validité ou avoir l'opportunité de l'améliorer; d'autre part, afin de connaître les méthodes permettant de documenter le rôle de la CI ou son impact dans le parcours étudiant.

3.3.1 Validité de contenu (interne)

Les premiers dispositifs ont mis à contribution l'expertise du chercheur (nous), afin de documenter la validité de contenu à l'instar de Katz (2007b) ou de Simard (2009). Trois dispositifs ont ainsi été appliqués.

Premièrement, la *Banque de questions* propose une formulation générique des items. Nous les avons adaptés aux disciplines¹⁰ de nos échantillons potentiels (sciences infirmières et travail social), de même qu'au contexte informationnel de l'UQO (ressources disponibles, environnement TIC, format des documents, etc.). Par exemple, prenons une question pour vérifier la connaissance de la ressource à privilégier pour repérer un livre. Dans la *Banque*, le titre du livre en exemple est un livre de collection générale et l'outil de recherche suggéré en choix de réponse correspond à l'environnement informationnel de l'Université de Montréal en 2007, un catalogue informatisé. Dans le QuizCI, l'item équivalent cite un livre disciplinaire et l'outil de recherche mentionné en choix de réponse est l'Outil de découverte du réseau de

¹⁰ Les diplômés en sciences infirmières ainsi que ceux en travail social peuvent être appelés à évoluer dans les mêmes milieux, particulièrement le système de santé et services sociaux du Québec, et auprès de clientèles pouvant présenter des problématiques similaires. Ainsi, les questions ont pu être formulées afin de sembler authentiques aux étudiants des deux programmes. Si bien que les items proposés aux étudiants des deux programmes d'études sont identiques.

l'Université du Québec (2009). Ainsi, l'ensemble des items a été révisé en fonction de l'environnement informationnel spécifique.

Deuxièmement, pour favoriser la validité de contenu du QuizCI, le type d'items proposés a été examiné. La recension des écrits avait déjà permis de connaître la principale critique adressée au test SAILS (Kent State University, 2015) en regard du type de ses items. En effet, ils sont à choix multiple. Ce type de données est réputé représenter plus difficilement une compétence et ses différents niveaux d'apprentissage (Allard et Gérin-Lajoie, 2011; Big6 Team, 2012). Lakhal, Frenette et Sévigny (2013) expliquent que :

« les méthodes d'évaluation traditionnelles (l'examen écrit et le test à choix multiples) visent à mesurer les niveaux cognitifs les plus bas [selon la Taxinomie de Bloom], soient la connaissance, la compréhension et l'application [...] alors que l'évaluation alternative (examen oral, travail en équipe, travaux pratiques, simulation, études de cas et projet) vise à mesurer des habiletés supérieures. » (p.134).

Pour contrer cette limite tout en conservant le format d'items à choix multiple, la deuxième démarche a consisté à varier la formulation des items afin de favoriser le recours aux différents niveaux cognitifs de l'étudiant (connaissances, compréhension, application, analyse, synthèse et évaluation) tout en ciblant les différentes dimensions de la CI. La taxonomie a été utilisée pour déterminer la difficulté attendue des items:

- Des questions de connaissance et de compréhension (A03, A04, A08, A10, A11, A12, B07, B10, B12, B13, B14, B15, B18, B20) visent à vérifier si l'étudiant connaît et comprend différentes notions de culture informationnelle. Assister à la formation et lire la documentation recommandée constituent les prérequis à la réussite de ces items.
- Des questions d'application et d'analyse (A05, A06, A09, B01, B03, B04, B08, B09, B16, B17, B19) présentent un problème informationnel où, pour trouver la réponse, l'étudiant doit parcourir le bon chemin et identifier la bonne réponse. Pour réussir ces items, il est suggéré d'avoir réalisé les exercices et de pouvoir transposer les apprentissages théoriques.
- Des questions de synthèse et d'évaluation font appel à des mises en situation (A01, A02, A07, B02, B05, B06, B11,) qui exigent l'évaluation itérative des sources, des contenus, des stratégies utilisées et des résultats de recherche.

Ainsi, les items du QuizCI visent à représenter les différents niveaux d'apprentissage. Le tableau de spécification rapporté dans le tableau 8 présente la répartition des items du QuizCI, selon les dimensions de la CI d'après la norme CRÉPUQ (2008) et par niveaux d'apprentissage de la taxonomie de Bloom révisée (Anderson, 1992).

Tableau 8

Tableau de spécification du QuizCI

| <u>Conceptualisation théorique du QuizCI</u> | | | | |
|--|-------------------------|-----------------------|--------------------------|------------------------|
| Niveau cognitif visé | Accéder à l'information | Évaluer l'information | Éthique informationnelle | Utiliser l'information |
| Connaissance | A12 A10 | A03 A04 | B15 | B14 |
| Compréhension | B07 B13 | A08 A11 | B18 B20 | B10 B12 |
| Application | B04 B09 | A05 A06 | B16 B17 | B11 |
| Analyse | B03 B08 | A09 | B19 | B01 |
| Synthèse | B02 B06 | A07 A02 | | |
| Évaluation | B05 | A01 | | |

Troisièmement, les items ont été révisés en fonction des analyses pressenties pour la présente étude. Le format des items de la *Banque* permettait parfois plus d'une bonne réponse. Par exemple, l'étudiant devait cocher tous les choix constituant une bonne solution. Ce format complique l'application d'un modèle de réponse à l'item. Comme il s'agit d'une première expérience pour l'auteure de la présente étude, Il a été privilégié d'utiliser la plus simple expression des modèles de réponse à l'item, celui à 1 paramètre ou modèle de Rasch. À cette

fin, les items ont été reformulés de façon à n'avoir qu'une seule bonne réponse et à collecter des données dichotomiques.

Ainsi, les trois dispositifs ont visé à ce que les items aient du sens dans le contexte disciplinaire et dans l'environnement informationnel de l'échantillon; qu'ils collectent des données de différents niveaux cognitifs ciblés par la formation documentaire *FD-UQO-SJ*; puis, que le format des données collectées permette de modéliser les données avec un modèle de Rasch.

3.3.2 Validité apparente

Parallèlement, nous avons examiné comment les étudiants comprennent ou reçoivent le QuizCI. Dans la *FD-UQO-SJ*, deux dispositifs récurrents colligent des preuves de validité apparente périodiquement auprès de l'échantillon visé. D'abord, à chaque collecte de données, fait suite une séance de rétroaction (tableau 9, activité #10). Elle permet de rencontrer et d'échanger avec les étudiants composant l'échantillon. Animé en grand groupe, ce dispositif contribue à sensibiliser les étudiants aux thèmes d'intérêt pour la validité. Cependant, étant donné que l'exercice a lieu pendant le temps de classe et souvent en compagnie de l'enseignant universitaire, nous avons anticipé un biais par désirabilité. Aussi, nous proposons également aux étudiants un formulaire anonyme pour rédiger leurs commentaires. Des modifications significatives ont découlé de ces deux dispositifs. Par exemple, nous avons clarifié certaines terminologies. Ou encore, nous avons procédé à certains arrimages entre les pratiques de formation documentaire et celles de l'enseignant disciplinaire. Les principaux résultats obtenus par les dispositifs de validité apparente en regard de la validité du QuizCI sont présentés en section 3.3.4.1.

3.3.3 Validité de contenu (externe)

Le troisième dispositif mis en place pour colliger des arguments sur la validité du QuizCI fait appel aux expertises pertinentes. À l'instar des textes retenus en recension des écrits, des experts de trois domaines ont été sollicités : un expert technologique; des experts bibliothécaires-formateurs; puis des experts disciplinaires, professeurs ou chargés du cours qui

accueillent la FD-UQO-SJ. Les résultats obtenus par les dispositifs de validité de contenu en regard de la validité de l'instrument sont présentés en section 3.3.4.2.

3.3.3.1. Expert technologique

Moodle est l'environnement numérique d'apprentissage (ENA) de l'UQO qui supporte le QuizCI. La personne-conseil au niveau de l'usage de Moodle, expert en technologie, a été consultée à deux reprises, d'abord à propos du design de l'instrument informatisé pour favoriser une l'utilisation efficace de l'ENA. Cette 1^{re} consultation a permis d'épurer le visuel et de retirer les mentions de texte pouvant nuire à la clarté de l'épreuve. La 2^e consultation a permis de sécuriser les données. Certains étudiants pouvaient les supprimer par mégarde. Ce risque a été amoindri avec l'aide de l'expert en technologie. Il a simplement programmé l'enregistrement des données à chaque question au lieu que ce soit fait à la fin de l'épreuve.

3.3.3.1. Experts bibliothécaires-formateurs

Quatre collègues bibliothécaires-formateurs (experts en CI) ont administré le QuizCI dans le cadre de leur formation documentaire. À la suite de quoi, ils ont été invités à commenter et contribuer à l'amélioration du questionnaire lors d'échanges verbaux. Un résumé de l'information obtenue de ces échanges est présenté en 3.3.4.2. Nous avons profité de ces échanges pour leur présenter le tableau de spécification. Trois des experts bibliothécaires ont validé le contenu du tableau de spécification du QuizCI. Cette réflexion transparaît dans le tableau de spécification retenu au final et présenté au tableau 8.

3.3.3.1. Experts disciplinaires

Ensuite, de 2009 à 2014 sept experts disciplinaires ont commenté le contenu du QuizCI. À ces occasions, des thèmes de discussion portant spécifiquement sur l'instrument ont été amenés. Ces échanges ont occasionné des ajustements entre autres sur les items retenus pour le QuizCI et leur formulation. Il a ainsi pu être discuté de la terminologie employée en formation et dans le QuizCI, des modes d'administration, durée et mode d'accès au QuizCI. De plus, en 2012, suite à la validation par les experts bibliothécaires, le tableau de spécification des items du QuizCI a été présenté à trois experts disciplinaires, deux professeurs en sciences infirmières et une enseignante universitaire en travail social. À ces occasions, nous avons pu vérifier la

compréhension commune du processus de mesure par rapport aux objectifs d'apprentissage de la formation documentaire, aux objectifs du cours et du programme universitaire. Un résumé de l'information obtenue lors de ces échanges est présenté en 3.3.4.2.

L'ensemble des démarches pour documenter la validité du QuizCI est résumé au tableau 9. Elles ont permis de colliger des arguments en faveur de la validité apparente aux yeux des étudiants, ainsi que de validité de contenu suite à l'examen par les experts consultés.

Tableau 9

Processus pour documenter la validité apparente et de contenu

| # | Session | Type de validité documentée | Démarches pour documenter la validité |
|--------|--------------|-----------------------------|---|
| 1 | Hiver 2009 | VC | Importer la traduction des items traduits du test SAILS et adapter la formulation des items d'un contexte spécifique |
| 2 | Hiver 2009 | VC | Varier la forme des items afin de favoriser le recours aux différents niveaux cognitifs de l'étudiant; méthode de validation : tableau de spécification |
| 3 – r* | Été 2009 | VC | Deux spécialistes du design informatisé d'instrument ont contribué au développement du QuizCI dans l'ENA Moodle |
| 4 | Automne 2009 | VA | Séance de rétroaction (tableau 2, activité 10) : échanges directs et ciblant de façon spécifique des thèmes d'intérêt pour la validité |
| 5 – r* | Automne 2009 | VA | Formulaire EFE (en annexe 1) destiné aux étudiants pour évaluer et rédiger leurs commentaires de façon anonyme. |
| 6 – r* | Automne 2010 | VC | Quatre experts en CI ont commenté et contribué à l'amélioration du questionnaire par un échange verbal. |
| 7 | Été 2011 | VC | Uniformiser les catégories de réponse pour n'avoir qu'un seul bon choix de réponses. |
| 8 | Automne 2011 | VC | Trois experts en CI ont validé le tableau de spécification du QuizCI |

(Tableau se continu)

| # | Session | Type de validité documentée | Démarches pour documenter la validité |
|--------|--------------|-----------------------------|---|
| 9 – r* | Automne 2011 | VC | Rencontre avec experts disciplinaires pour rétroaction verbale, mais ciblant spécifiquement des éléments de validité |
| 10 | Hiver 2012 | VC | Appariement par un expert disciplinaire du contenu de l'instrument, de la formation documentaire et des objectifs du programme d'études |

Légende : r* = Activité récurrente à chaque FD ; VA = validité apparente ; VC = validité de contenu.

3.3.4. Résultats sommaires du processus de validation préalable du QuizCI

Les résultats du processus préalable de validation ont permis de collecter de l'information pouvant être mise en perspective avec celle des dispositifs de validité conceptuelle. Les principaux éléments d'information qui en ressortent sont ici présentés.

3.3.4.1. Résultats obtenus des dispositifs de validation apparente

Depuis 2009, chaque groupe d'étudiants ayant complété le QuizCI est invité à remplir un questionnaire EFE (Activité # 5 - r, tableau 9). Notre compilation des résultats de l'EFE (en annexe 1) de l'automne 2014 fait ressortir trois commentaires ayant obtenu un haut niveau d'accord de la part des étudiants, ce qui est conforme au point de vue des étudiants souvent émis verbalement :

- Cette formation est essentielle à mon parcours universitaire.
- Je compte utiliser régulièrement les outils de recherche.
- Il s'agit d'un bon complément d'information utile à mon programme universitaire.

Tandis que les trois commentaires libres (rédigés dans la section commentaires) les plus fréquents sont :

- Le QuizCI est trop difficile.
- La formation va trop vite.
- Il y a trop d'informations.

Ces résultats montrent que les étudiants accordent une importance à la CI, mais qu'ils ont l'impression que l'effort requis est trop élevé.

3.3.4.2. Résultats obtenus des dispositifs de validation de contenu

Les experts bibliothécaires-formateurs ont permis deux constats en regard de la validité de contenu du QuizCI. Le consensus fait autour du tableau de spécification nous informe que, théoriquement, ils reconnaissent les qualités du QuizCI et que son contenu représente bien le construit théorique présenté dans le tableau de spécification. Toutefois, en pratique lors de l'usage du QuizCI dans leur formation documentaire, les bibliothécaires-formateurs ont tous réagi de la même façon : ils ont simplifié et allégé le niveau de difficulté de l'épreuve (automne 2013 et hiver 2014). Ces modifications ont toujours été faites sans questionnement quant au bien-fondé de la démarche, mais avec de bonnes intentions. Par conséquent, les données ainsi collectées ne correspondent plus au QuizCI ayant fait l'objet d'un processus de validation et ne peuvent donc pas être retenues dans l'échantillon analysé dans le cadre de la présente étude. À l'instar de McMillen et Deitering (2007), les données 2013 et 2014 ont été retirées de notre échantillon.

Deux experts disciplinaires, professeurs encadreurs des cours concernés, ont respectivement jugé que le tableau de spécification du QuizCI est adéquat en regard des objectifs du cours et de leur programme universitaire (SOI et TSO). Ensuite, des échanges avec sept enseignants universitaires nous informent qu'ils sont unanimes sur l'importance de la CI dans le parcours universitaire. Ils estiment que la forme du questionnaire est adéquate. Puis, ils jugent l'usage de pratique d'évaluation nécessaire, d'une part pour que l'évaluation de l'ensemble du cours reflète cette portion de matière enseignée dans le cadre du cours, d'autre part, pour que les étudiants accordent l'importance requise à cette matière moins disciplinaire et souvent mal acceptée.

Enfin, la version du QuizCI composée de 32 items dont la validité a été documentée et éprouvée à l'aide de 10 dispositifs constitue *LE QuizCI* ayant servi à collecter les données destinées à documenter la validité conceptuelle de l'instrument dans la présente étude.

3.4. Participants

Les participants à la présente étude sont des étudiants inscrits aux cours universitaires ciblés par la FD. Ainsi, l'échantillon de données a été collecté auprès d'étudiants à l'Université du Québec en Outaouais campus Saint-Jérôme. Il s'agit d'un échantillon systématique, où tous les étudiants des groupes-cours ciblés par la formation documentaire ont été retenus pour l'étude. L'administration du QuizCI s'est effectuée auprès de 16 groupes-cours (voir tableau 10), 15 en sciences infirmières (ssa1333) et un en travail social (tso1003) tel que présenté dans le tableau 10. Au départ, 523 étudiants ont complété le QuizCI. De ceux-ci, 497 étudiants ont accordé leur consentement à ce que les données de leur QuizCI contribuent à la présente étude selon le certificat éthique CPER-14-093-D (2013, 2014). Vingt-six étudiants n'ont pas souhaité y participer. De ces 497 données collectées, 28 étudiants n'ont complété qu'une moitié de Quiz. En effet, dans le cadre du cours ssa1333, la FD est scindée en 2 parties (tableau 2, activités #4 et #7). Aussi, le QuizCI est proposé aux étudiants en deux moitiés, suite à chacune des séances de la FD. Ces 28 étudiants n'ont complété qu'une moitié; généralement ayant abandonné le cours entre les 2 activités. Ces patrons incomplets ont été retirés de l'échantillon. Au final, l'effectif des participants s'élève à n=469 patrons de réponses complets, sans aucune donnée manquante.

La collecte de données s'est déroulée sur trois sessions universitaires (voir tableau 10). À l'automne 2012, six groupes en sciences infirmières ont permis de collecter des données auprès de 186 étudiants (SOI), alors qu'un groupe en travail social (TSO) a permis de collecter des données auprès de 34 étudiants. À l'hiver 2013, deux groupes en sciences infirmières ont permis d'ajouter 69 étudiants à l'échantillon. Puis, à l'automne 2014, sept groupes en sciences infirmières ont permis au QuizCI de rejoindre 219 étudiants.

Tableau 10

Données collectées et échantillon retenus

| Groupe (GR) | Échantillon (<i>n</i>) | Moyenne des scores | Écart-type des scores |
|---------------------|--------------------------|--------------------|-----------------------|
| Automne 2012 | | | |
| SOI ₀₁ | 29 | 21,28 | 3,422 |
| SOI ₀₂ | 24 | 20,75 | 2,801 |
| SOI ₀₃ | 25 | 21,76 | 3,032 |
| SOI ₀₄ | 15 | 21,73 | 3,634 |
| SOI ₀₅ | 30 | 19,70 | 4,145 |
| SOI ₀₆ | 29 | 21,45 | 3,146 |
| TSO ₀₇ | 34 | 23,38 | 3,635 |
| Hiver 2013 | | | |
| SOI ₀₈ | 35 | 22,74 | 2,984 |
| SOI ₀₉ | 29 | 22,07 | 2,448 |
| Automne 2014 | | | |
| SOI ₁₀ | 34 | 24,09 | 3,127 |
| SOI ₁₁ | 33 | 22,39 | 2,872 |
| SOI ₁₂ | 40 | 23,98 | 3,254 |
| SOI ₁₃ | 20 | 24,35 | 3,048 |
| SOI ₁₄ | 38 | 23,95 | 2,265 |
| SOI ₁₅ | 22 | 23,68 | 3,617 |
| SOI ₁₆ | 32 | 22,22 | 3,045 |
| Échantillon complet | 469 | 22,55 | 3,371 |

Enfin, mentionnons que le QuizCI a été administré à des étudiants de plus de 18 ans, admis à l'université, qui ont tous satisfait aux exigences en langue française d'admission à l'université, soit l'examen uniforme collégial (Ministère de l'Éducation du Sport et du Loisir, 2001). Cette caractéristique des participants prend son sens du fait que les items du QuizCI exigent une habileté en lecture de niveau universitaire.

3.5. Les données

Les données sont donc constituées par les réponses de 469 étudiants aux 32 items du Quiz CI. Chaque patron de réponses retenu dans l'échantillon a été anonymisé. L'identité de l'étudiant a été remplacée par un identifiant à cinq chiffres; les deux premiers retracent le groupe; les trois derniers chiffres de l'identifiant est unique pour chaque patron de réponse. Pour réaliser les analyses, deux logiciels ont été utilisés, SPSS version 22 et Winsteps (2007). Pour saisir les données dans chaque logiciel, les données ont été recodées de façon à présenter les données dichotomisées (pour les raisons évoquées en section 3.3.). Pour le logiciel SPSS, les données ont été recodées tel que 0 = échec, 1 = succès. Tandis que pour le logiciel Winsteps, elles ont été codées tel que 1 = échec, 2 = succès, car le zéro a une autre fonction pour Winsteps.

3.6. Méthodologies pour documenter la validité conceptuelle

Pour documenter la validité conceptuelle du QuizCI, les dispositifs pressentis sont des analyses quantitatives. Elles sont réalisées pour produire des preuves de validité conceptuelle telle que décrite par Laveault et Grégoire (2002).

Pour ce faire, des éléments de preuves selon les trois domaines de la validité précisés par Borsboom, Mellenbergh et van Heerden (2004) seront amenés. Pour chaque dispositif, l'analyse elle-même sera présentée, son apport, la règle de décision retenue, puis le résultat obtenu pour notre échantillon. Pour vérifier si l'instrument réagit en fonction de la CI, trois paramètres seront calculés : l'indice de simplicité, l'indice de discrimination et le coefficient alpha de Cronbach. Pour documenter comment le QuizCI représente la CI dans sa nature et sa structure conceptuelle, une analyse factorielle exploratoire sera réalisée. Pour examiner si et comment les données collectées avec le QuizCI permettent de mesurer la CI, un modèle de Rasch sera appliqué aux données avec le logiciel Winsteps. Un processus d'ajustement sera réalisé. Le modèle final proposé sera examiné afin de discuter de validité.

Enfin, la méthodologie ici énoncée propose de réaliser certaines analyses qui fournissent de l'information similaire par des processus différents. Par exemple, l'indice de simplicité et la représentation de l'échelle de Rasch permettent d'apprécier la difficulté de l'item. L'analyse factorielle exploratoire et l'analyse des vecteurs de dimensions principales fournissent des représentations de la dimensionnalité de l'instrument. Nous recourons à plus d'un point de vue,

car à ce moment-ci de la démarche, il nous est encore impossible de retenir ou de rejeter l'une ou l'autre des procédures, de plus que la preuve de validité conceptuelle repose sur une accumulation d'information/preuve dans un même sens.

Chapitre 4 : résultats

Ce 4^e chapitre présente les résultats d'analyses quantitatives réalisées pour documenter la validité conceptuelle du QuizCI. Les informations obtenues à l'aide de chacun des dispositifs visent à dégager des arguments ou des preuves de validité du QuizCI. Ces analyses sont réalisées dans le but de documenter les trois domaines de la validité. Premièrement, trois analyses issues de la théorie classique des tests sont réalisées afin de documenter comment les scores au QuizCI varient en fonction de la CI de l'étudiant. Deuxièmement, un processus d'analyse de la dimensionnalité du QuizCI permet de documenter la nature et la structure du QuizCI. Troisièmement, les méthodes de la TRI permettent de documenter les liens entre la CI des étudiants et leur score au QuizCI. Les résultats obtenus seront mis à contribution au 5^e chapitre pour discuter de la validité du QuizCI. Une analyse descriptive des données est d'abord présentée.

4.1. Analyse descriptive des données

L'échantillon compte 469 patrons pour 32 items, sans données manquantes. Avant même de mettre en œuvre les dispositifs retenus pour documenter la validité conceptuelle, nous présentons ici quelques analyses pour cerner la qualité de l'échantillon disponible. D'abord une analyse de la distribution des scores totaux vérifie la normalité de la distribution des scores du QuizCI. Ensuite, la taille de l'échantillon et l'homogénéité de la variance sont examinées.

4.1.1 Distribution des scores totaux

L'histogramme de la distribution des scores totaux (figure 2) montre que le score total le plus faible est de 10/32 (31%) et le plus fort est de 31/32 (97%). La moyenne des scores totaux est de 22,47/32 (70%). Tandis que l'écart-type est de 3,371 points. Le score médian et le mode sont de 23/32 (71,9%). La courbe présente une légère asymétrie négative.

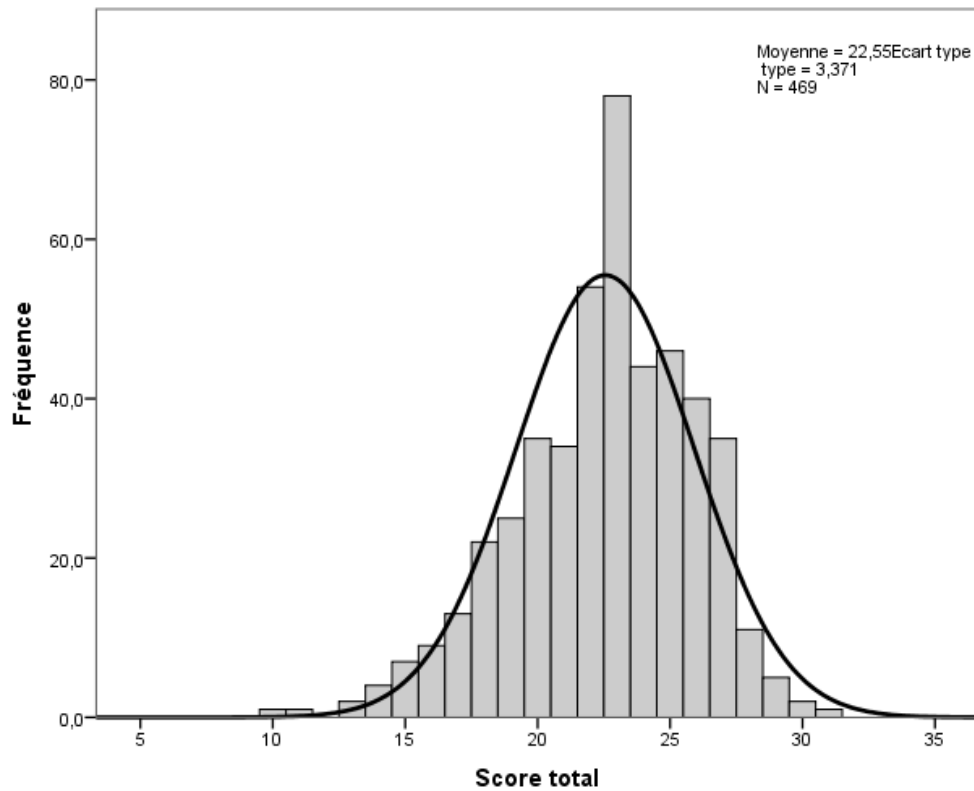


Figure 2. Histogramme des scores totaux du QuizCI.

Pour être considérée comme une courbe normale, le coefficient d'asymétrie doit être compris entre -1 et 1, et le coefficient d'aplatissement entre -1,5 et 1,5 (Leech, Barrett et Morgan, 2015). Pour cet échantillon, le coefficient d'asymétrie est de -0,437 et le coefficient d'aplatissement est de -0,659. Les scores totaux au QuizCI de cet échantillon peuvent donc être considérés comme normalement distribués. Il s'agit d'un postulat pour réaliser certaines analyses comme l'analyse factorielle exploratoire.

4.1.2. Taille d'échantillon

Certaines de nos analyses ont un postulat en regard de la taille d'échantillon requise, entre autres lors de l'application d'un modèle de Rasch, ou pour réaliser une l'analyse factorielle. Pour l'application d'un modèle de Rasch, les études recensées se réfèrent à la règle générale de 5 personnes par items [5 : 1] (Bond et Fox, 2007). Pour l'analyse factorielle exploratoire, Bourque, Poulin et Cleaver (2006) suggèrent un ratio de [10 : 1].

Notre échantillon compte 469 étudiants pour 32 items, soit un ratio de 14,65 patrons de réponses par item. Il répond donc à ces règles générales. Par ailleurs, Tabaschnick et Fidell (2013) recommandent de rester vigilant à propos de cette exigence. Elle peut varier spécifiquement pour chaque instrument.

4.1.3. Homogénéité de la variance

Lors de la représentation de la dispersion des données au fur et à mesure de leur obtention, le positionnement de la droite des moindres carrés pour l'ensemble des données montre une légère pente ascendante illustrée par la droite dans la figure 3. Ce constat nous a mené à nous questionner quant à l'existence possible d'un biais lié à la collecte de données par groupe. Ce biais pourrait par exemple être lié à la prestation du formateur qui adapte son contenu au fil de ses prestations ou bien à une préparation accrue des étudiants à passer le test au fil du temps.

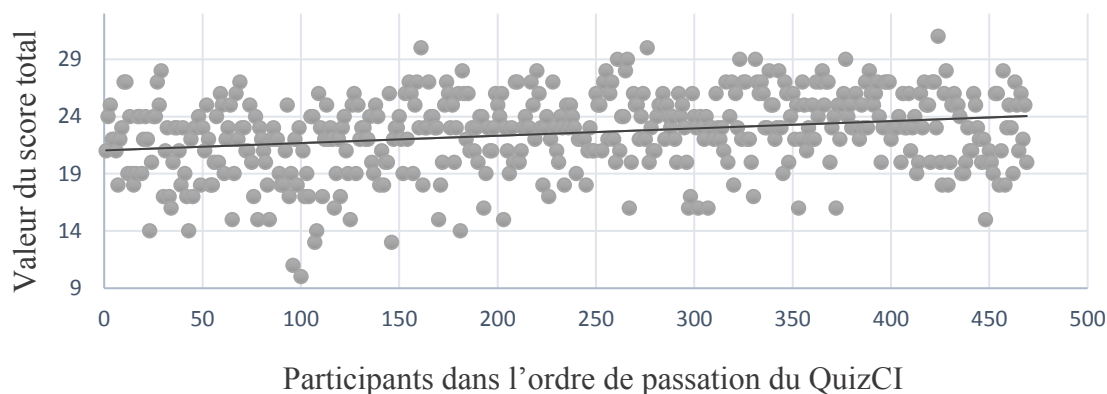


Figure 3. Dispersion des scores des participants

Les données ont été collectées auprès de 16 groupes-cours et au fil de quatre sessions universitaires (tableau 10). Or, la moyenne respective des groupes varie puisque la moyenne du groupe le plus faible (# 5) est de 19,70/32 (62%) tandis que celle du groupe le plus fort (#13) est de 24,35/32 (76%). Nous avons cherché à vérifier si nous pouvons quand même considérer les groupes comme provenant de la même population (Tabaschnick et Fidell, 2013). Nous avons ainsi produit un diagramme présentant toutes les moyennes des groupes avec leur intervalle de confiance à 95% (figure 4).

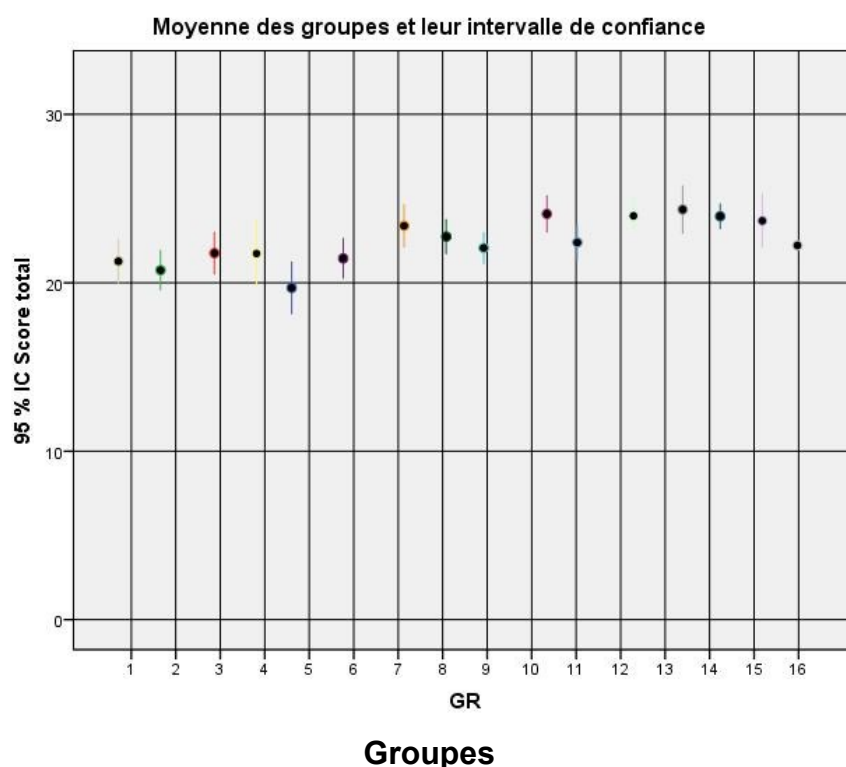


Figure 4 : Moyenne des groupes avec barres d'erreur.

La figure 4 met en évidence les différences entre certaines moyennes, par exemple entre les groupes 5 et 13, pour lesquels les intervalles sont disjoints. L'existence de différences est confirmée par une ANOVA ($F(15, 453) = 1,58$; $p = 0,00 < 0,05$). Le test post hoc de Bonferroni confirme que les deux moyennes les plus faibles ($gr_5 : 19,70$ et $gr_2 : 20,75$) diffèrent de certaines autres moyennes, particulièrement des groupes où la moyenne est élevée. Les moyennes restent quand même assez stables, nous avons décidé de garder tous les groupes dans notre échantillon, considérant que notre démarche de validation n'en souffrirait pas.

En somme, les analyses préalables ont permis de savoir que

- les données collectées avec le QuizCI se distribuent selon une courbe considérée normale;
- la taille de l'échantillon est théoriquement suffisante pour réaliser les analyses pressenties;
- les moyennes de deux des groupes sont plus faibles que les autres.

4.2. Analyses et résultats issus de la théorie classique des tests

Le premier domaine de la validité conceptuelle examine comment les scores au QuizCI fluctuent en fonction de la variation de la CI chez l'étudiant. Pour la présente étude, cet examen repose sur trois analyses issues de la théorie classique des tests (TCT) : l'analyse de la difficulté des items à l'aide de l'indice de simplicité; puis, l'analyse de la capacité de l'instrument à discriminer les étudiants en fonction de leur CI; l'analyse de la cohérence interne des items à l'aide du coefficient alpha de Cronbach. Les analyses de la TCT n'exigent aucun postulat.

4.2.1. Difficulté de l'item

La difficulté des items est ici examinée à l'aide de l'indice de simplicité (*IS*). Cet indice correspond au pourcentage des étudiants qui ont répondu correctement à la question. Par exemple, une question réussie par seulement 23% des répondants présente un *IS* de 0,23. Pour en interpréter le résultat, Mery, Newby et Peng (2011) présentent une règle de décision subdivisée en huit catégories. Étant donné le caractère exploratoire de notre recherche, nous avons retenu une règle de décision plus sommaire à trois catégories proposée par Zubairi, Lide et Kassim (2006) : trop facile ($IS \leq 0,32$), adéquat ($0,33 \leq IS \leq 0,68$) et trop difficile ($IS > 0,69$).

Le tableau 11 présente les *IS* des items du QuizCI. Neuf items (28% de l'effectif) présentent une difficulté de niveau adéquat pour l'échantillon, se distribuant dans l'intervalle de $0,33 \leq IS \leq 0,68$. Par ailleurs, 22 items (69% de l'effectif) présentent un indice de simplicité supérieur à 0,69 et se classent comme étant trop faciles pour les sujets de l'échantillon. Un seul item du QuizCI est ici identifié comme trop difficile, B11 présente un *IS* de 0,23.

Tableau 11

Indice de simplicité des items du QuizCI

| Règle de décision | Items |
|--|--|
| Items trop difficiles $0,32 \leq IS$ | B11 |
| Items adéquats $0,33 \leq IS \leq 0,68$ | A12, B20, B08, B13, B07, B01, A06, B04, A11 |
| Items trop faciles $IS > 0,69$ | A04, A10, B15, B03, A08, B12, B06, A03, B17, A07, B02, A01, B09, B10, B18, A05, B14, B16, A09, B19, B05, A02 |

Source: Mery, Y., Newby, J., & Peng, K. (2011). Assessing the reliability and validity of locally developed information literacy test items. *Reference services review*, 39(1), p.106.

4.2.2. Discriminer les étudiants en fonction de leur CI

Pour vérifier la capacité des items du QuizCI à discriminer les étudiants en fonction de leur CI, l'indice de discrimination (D) est utilisé. Il présente la différence entre le résultat moyen des étudiants ayant obtenu un résultat plus élevé (Re) à l'épreuve et celui des étudiants ayant obtenu un résultat plus faible (Rf). Pour le calculer, le résultat (R), soit les scores des étudiants, est ordonné et scindé en quartile.

Pour notre échantillon de 469 sujets, les quartiles sont respectivement composés de 117 étudiants. Lorsque les scores sont ordonnés du plus élevé au plus faible, le 1^{er} quartile (Q_1) regroupe les 117 étudiants ayant le plus bas score, soit tous les scores inférieurs à 20 sur 32 ($R_{Q_1} \leq 20$). Le 2^e quartile regroupe 117 étudiants suivants, soit occupant les rangs de 118 au 234. Les scores concernés sont de 21 ou 22 sur 32 (R_{Q_2} 21-22). Le 3^e quartile rassemble les 117 étudiants suivants, occupant les rangs de 235 à 351. Les scores totaux concernés s'étendent de 23 à 25 sur 32 (R_{Q_3} 23-25). Puis, le 4^e quartile (Q_4) compte les 118 étudiants restants, dont les scores sont de 26 ou plus sur 32 ($R_{Q_4} \geq 26$).

Pour chaque item, l'indice de discrimination de l'item représente la différence entre le résultat moyen du quartile plus élevé (R_e) et celui du quartile faible (R_f). Par exemple, l'item A02 a été réussi par 113 des 118 étudiants de R_{Q4} . Le score moyen à cet item pour le quartile le plus élevé (R_e) est de 99%. Tandis que le résultat moyen pour R_{Q1} , quartile des scores les plus faibles est de (R_f : 89%). La différence 10% aussi l'indice de discrimination pour l'item A02 est 0,10.

Pour interpréter les indices de discrimination, Mery, Newby et Peng (2011), ainsi que Zubairi, Lide et Kassim (2006) décrivent l'usage d'une règle de décision en cinq catégories (tableau 7). En raison du caractère exploratoire de la présente étude, nous retenons seulement trois catégories selon la même logique que celle appliquée à l'IS. Lorsque D est égal ou supérieur à 0,20; les items discriminent de façon satisfaisante. Les items dont l'indice de discrimination est insatisfaisant présentent un indice de discrimination située par l'intervalle de 0,00 à 0,19. Ces items doivent être révisés. Enfin, les items présentant un indice de discrimination inférieur à zéro, soit négatif, nuisent à la capacité de discrimination de l'instrument et doivent être éliminés.

Le résultat du calcul de l'indice de discrimination des items du QuizCI est présenté au tableau 12. Il ressort que vingt (20) items présentent une capacité de discrimination supérieure au seuil de $D \geq 0,20$ et discriminent suffisamment la CI des étudiants. Tandis que dix items présentent une capacité de discrimination dans l'intervalle de 0,00 à 0,19 et ne discriminent pas suffisamment la CI des étudiants. Ces items doivent être révisés. Enfin, deux items obtiennent un indice de discrimination dont la valeur est négative. Ces items nuisent à la capacité de discrimination de l'instrument et doivent être éliminés.

Tableau 12

Capacité de discrimination de l'item du QuizCI en fonction de la CI de l'étudiant

| Règles de décision | Items |
|---|--|
| Item discrimine adéquatement en fonction de la compétence informationnelle des étudiants D supérieur 0,20 La capacité de discrimination de l'item en fonction de la compétence informationnelle de l'étudiant en insuffisant. | B13, A11, A12, B08, B07, A10, B15, B06, B09, B12, B10, A06, B01, B02, A07, B04, A01, A08, B16, B17 |
| D compris entre 0,00 - 0,19 Item nuit à la mesure est négative D inférieur 0,00 | A05, B14, B03, B05, A09, B19, A04, B18, B11, A02 A03, B20 |

Source: Mery, Y., Newby, J., & Peng, K. (2011). Assessing the reliability and validity of locally developed information literacy test items. *Reference services review*, 39(1), p.108.

4.2.3. Cohérence interne

Documenter la cohérence interne du QuizCI permet de savoir à quel point les items travaillent ensemble à mesurer un même trait. La cohérence interne est généralement représentée par le coefficient alpha de Cronbach, α . Il est calculé pour un ensemble d'items présumés représenter une mesure unidimensionnelle. Le coefficient alpha de Cronbach varie entre 0 et 1. Plus sa valeur s'approche de 1, plus ce résultat signifie que les items travaillent ensemble pour la mesure. Rappelons la règle de décision proposée par Cronbach en 1951 (L.J. Cronbach, 1951) (L.J. Cronbach, 1951) (L.J. Cronbach, 1951) (L.J. Cronbach, 1951) :

α = inférieur à - 0,4 insatisfaisant

α = [0,4 - 0,6] faible

α = [0,6 - 0,8] satisfaisant

α = supérieur à 0,8 - excellent

Le coefficient alpha de Cronbach de tous les items du QuizCi présente une cohérence interne faible de 0,501. Pour investiguer quels items sont plus problématiques, nous avons consulté l'alpha de Cronbach *en cas de suppression de l'item proposé par SPSS*. Pour chaque

item, il correspond au coefficient alpha de Cronbach pour tous les autres items de l'épreuve. Les items dont l'alpha est supérieur *sans l'item* nuisent à la cohérence interne.

Nous avons identifié six (6) items nuisant à la cohérence interne de la mesure du QuizCI. En ordre du plus nuisible au moins nuisible se trouvent les items B20, A03, A04, B18, B11 et B03. Le retrait de tous ces items permet d'obtenir le meilleur alpha de Cronbach calculé avec cet échantillon, soit $\alpha = 0,636$; valeur qui se classe comme satisfaisante (Cronbach, 1951) . Il s'agit d'une preuve que vingt-six items sur les 32 travaillent ensemble pour mesurer la CI¹¹.

En somme, les analyses de la TCT ont permis de documenter la validité conceptuelle du QuizCI pour examiner comment les scores au QuizCI fluctuent en fonction de la variation de la CI chez l'étudiant. Pour ce faire, différentes analyses ont été réalisées : une analyse de la difficulté des items à l'aide de l'indice de simplicité; un examen de la capacité des items à discriminer les étudiants en fonction de leur CI; ainsi qu'une analyse de la cohérence interne de la mesure à l'aide du coefficient alpha de Cronbach.

Le résultat de l'analyse de la difficulté des items suggère que l'épreuve est de façon générale trop facile pour les étudiants de notre échantillon. Près de 70% des items (22) se catégorisent comme trop faciles. Le résultat de l'analyse de la capacité de discrimination des items identifie 10 items discriminants de façon insuffisante, tous ces items étant identifiés comme trop faciles. Notons que deux items se sont catégorisés comme nuisant à la capacité de discrimination du QuizCI, l'un est trop facile et l'autre de difficulté adéquate. Tandis que le résultat de la cohérence interne a révélé que 26 des 32 items travaillent ensemble pour mesurer, alors que six items nuisent à la cohérence interne de l'instrument. La cohérence interne du QuizCI est *satisfaisante* avec un coefficient alpha de Cronbach pour la solution optimale de 0,636.

¹¹. Nous avons également examiné si les items du QuizCI subdivisés par facteur identifié par l'AFE (section 4.3.1) présentent une cohérence interne supérieure, même si cela était peu probable étant donné le petit nombre d'items dans chaque facteur. Le résultat pour chacun des facteurs est inférieur à 0,5 : $\alpha_{f1} = 0,318$; $\alpha_{f2} = 0,465$. La cohérence interne des items n'est donc pas meilleure par facteur. Aussi, nous conservons une mesure unidimensionnelle plus conforme à nos attentes.

Le tableau 13 présente une synthèse des informations permettant d'examiner comment le QuizCI réagit à la CI.

- Huit items discriminent adéquatement les étudiants en fonction de leur CI, contribuent à la cohérence interne du QuizCI et présentent un niveau de difficulté adéquat pour l'échantillon de sujets.
- Douze items discriminent adéquatement les étudiants en fonction de leur CI, contribuent à la cohérence interne du QuizCI et présentent un niveau de difficulté facile pour l'échantillon de sujets.
- Six items contribuent à la cohérence interne du QuizCI, mais sont faciles et pour cette raison discriminent de façon insuffisante les étudiants en fonction de leur CI.
- Six items nuisent à la cohérence interne du QuizCI. De ceux-ci, quatre discriminent de façon insuffisante les étudiants en fonction de leur CI, tandis que deux nuisent à la capacité de discrimination de l'instrument.

Tableau 13

Synthèse des informations sur comment le QuizCI réagit en fonction de la CI.

| Items | <i>D</i> | α | <i>IS</i> |
|---|-----------------------|---------------------|-----------|
| A06, A11, A12, B01, B04, B07, B08, B13 | Adéquat | Contribue | Adéquat |
| A01, A07, A08, A10, B02, B06, B09, B10, B12, B15, B16, B17 | Adéquat | Contribue | Facile |
| A02, A05, A09, B05, B14, B19 | Insuffisant | Contribue | Facile |
| A04, B03, B18 | Insuffisant | α mieux sans | Facile |
| B11 | Insuffisant | α mieux sans | Difficile |
| A03 | Item nuit à la mesure | α mieux sans | Facile |
| B20 | Item nuit à la mesure | α mieux sans | Adéquat |

4.3. Analyses et résultats pour documenter le concept de CI

Le second domaine de la validité conceptuelle vise à documenter la conceptualisation de la CI telle que représentée par les données du QuizCI. Pour la présente étude, le dispositif utilisé pour produire de l'information sur cette question est l'analyse factorielle exploratoire (AFE) réalisée avec le logiciel SPSS. Cette analyse s'effectue avec l'échantillon de 469 patrons de réponses d'étudiants ayant complété le QuizCI de 32 items. Le ratio est donc d'un item pour 14,65 patrons de réponses, ce qui est supérieur au ratio recommandé pour l'AFE de 1 :10 (Bourque et al., 2006). Les données sont dichotomisées en bonne ou mauvaise réponse, caractéristique déterminante dans les choix méthodologiques lors d'une AFE.

4.3.1. Analyses préalables

Certaines analyses préalables permettent de vérifier s'il est adéquat de réaliser une analyse factorielle exploratoire (AFE) sur cet échantillon de données. Il s'agit d'examiner le caractère factorisable des items. Pour documenter le caractère factorisable des données collectées avec les items du QuizCI en termes de postulat, nous vérifions la mesure de Kaiser-Meyer-Olkin (KMO) et le test de sphéricité de Bartlett.

La mesure KMO est un indice d'adéquation de la solution factorielle. Elle vérifie l'adéquation de l'échantillon en vue d'une réduction des dimensions et donne un aperçu global de la qualité des corrélations inter-items. La mesure KMO recommandée se situe entre 0 et 1 ; où plus la valeur s'approche de 1, plus la mesure est bonne (Durand, 2003). Notre échantillon obtient une mesure KMO de 0,810. Ce résultat suggère deux arguments de validité : il indique que notre échantillon constitue un ensemble suffisamment cohérent pour constituer une mesure adéquate d'un concept. Ensuite, il suggère qu'il existe une solution factorielle statistiquement acceptable représentant la structure dimensionnelle du construit observé.

Le test de Bartlett vérifie l'hypothèse nulle selon laquelle toutes les corrélations seraient égales à zéro. Pour une factorisation dans le cadre d'une AFE, une certaine corrélation est nécessaire. Aussi, le test doit être significatif. Le résultat du test de sphéricité de Bartlett pour l'échantillon du QuizCI est significatif, $p = 0,00$; ce qui est inférieur au seuil de 0,05.

L'hypothèse nulle est rejetée. Les corrélations ne sont donc pas toutes égales à zéro. Aussi, les items entretiennent certains liens entre eux; ce qui est propice à la factorisation.

Ainsi, les résultats aux analyses préalables indiquent que notre échantillon de données est propice à la réalisation d'une analyse factorielle exploratoire.

4.3.2. Analyse factorielle exploratoire

L'analyse factorielle exploratoire est réalisée par la *réduction de dimensions* de SPSS. Dans cette procédure, le chercheur doit effectuer certains choix déterminants pour le résultat. Il doit sélectionner la méthode d'extraction, la rotation de la solution et le nombre de facteurs retenus dans la solution factorielle. Nous avons effectué les choix suivants :

1. Comme nous l'avons vérifié préalablement à l'aide du coefficient d'asymétrie et d'aplatissement de la courbe, la distribution des items est normale et justifie l'usage de la méthode d'extraction du maximum de vraisemblance (maximum likelihood) (Durand, 2003; Tabaschnick, 2013).
2. Les données dichotomiques invitent à l'application de la rotation oblique *Oblimin*, visant à minimiser le nombre d'items attribués à un facteur donné en maximisant la variance intra-facteurs, c'est-à-dire en favorisant l'écart entre les coefficients de saturation factorielle élevés et faibles (Tabachnick et Fidell, 2013). Cette rotation est appropriée aux structures unidimensionnelles et multidimensionnelles (Bourque, 2006). Pour cette raison, elle est adaptée à l'étude de trait latent en sciences sociales, car elle accepte une certaine corrélation entre les facteurs (Durand, 2003).
3. Ensuite, le tracé d'effondrement (figure 5) des « valeurs propres » permet d'identifier le nombre de facteurs à retenir dans la solution factorielle. Les facteurs formant une pente accentuée expliquent la majeure partie de la variance. Les facteurs s'alignant horizontalement, eux, apportent peu de nouvelles informations.

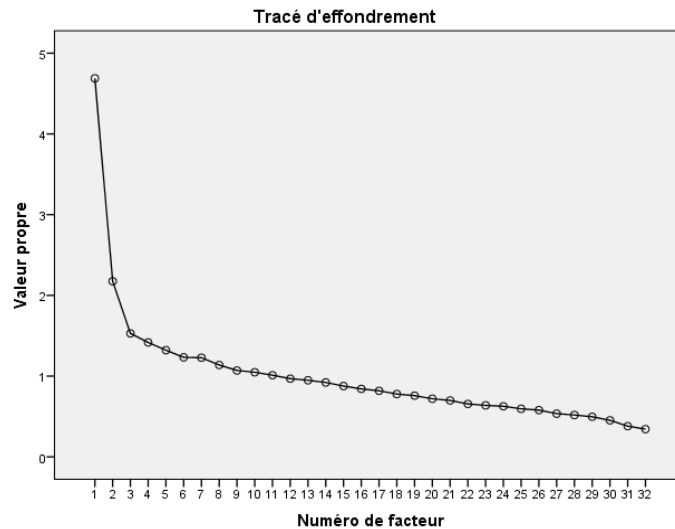


Figure 5. Tracé d'effondrement des valeurs propres des données collectées avec le QuizCI.

Le tracé d'effondrement de l'échantillon du QuizCI suggère de retenir une solution factorielle à deux facteurs, puisque la courbe devient quasi horizontale après le second facteur.

Ainsi, à la lumière de ces choix, la bonne démarche de réduction de dimensions peut être réalisée, soit en utilisant la méthode d'extraction du maximum de vraisemblance, en appliquant la rotation oblique Oblimin et en imposant une limite de deux facteurs à la solution. La solution factorielle retenue est présentée au tableau 15.

Auparavant, parmi les résultats de l'analyse, le tableau 14 présente la qualité de représentation des items dans la factorisation. La qualité de la représentation indique l'appartenance de chaque item à la covariance de l'ensemble des items en tant que facteur. Durand (2003) considère que la qualité de la représentation doit être supérieure à 0,20 pour justifier le maintien de la variable dans l'analyse.

Le tableau 14 identifie les items A02, A08 et B11 comme présentant une qualité de représentation inférieure à 0,20 à interpréter avec prudence. De façon cohérente, dans la solution factorielle présentée au tableau 15, ces itemsaturent faiblement sur leur facteur respectif. A02 est l'item qui sature le moins sur le 1^{er} facteur. A08 est l'item qui sature le moins sur le 2^e facteur. Puis, B11 n'atteint pas un seuil de saturation suffisant sur aucun des 2 facteurs pour être retenu dans la solution factorielle.

Tableau 14

Qualité de représentation des facteurs par les données de l'échantillon.

| Item | Qualité de représentation de l'item |
|-------------|-------------------------------------|
| Facteur 1 | |
| A01 | 0,287 |
| A02* | 0,108 |
| A03 | 0,360 |
| A04 | 0,317 |
| A05 | 0,207 |
| A06 | 0,200 |
| A07 | 0,257 |
| A08* | 0,198 |
| A09 | 0,273 |
| A10 | 0,347 |
| A11 | 0,293 |
| A12 | 0,534 |
| B01 | 0,229 |
| B02 | 0,286 |
| B03 | 0,208 |
| B04 | 0,247 |
| Facteur 2 | |
| B05 | 0,387 |
| B06 | 0,326 |
| B07 | 0,562 |
| B08 | 0,229 |
| B09 | 0,255 |
| B10 | 0,333 |
| B11* | 0,006 |
| B12 | 0,228 |
| B13 | 0,384 |
| B14 | 0,265 |
| B15 | 0,298 |
| B16 | 0,275 |
| B17 | 0,214 |
| B18 | 0,254 |
| B19 | 0,214 |
| B20 | 0,320 |

Légende : Les items en gras marqués d'un astérisque sont les items atteignant une saturation inférieure à 0,2; seuil inférieur en contexte exploratoire.

La solution factorielle elle-même est présentée au tableau 15. Il est admis que pour retenir un item comme contribuant à la variance d'un facteur, une saturation supérieure à 0,3 sur un

seul facteur est recherchée (Tabachnick et Fidell, 2013). La solution factorielle retenue comme la meilleure montre qu'il y a un problème d'adéquation pour plusieurs items. Du facteur 1, les items B18, B14, B12 et A08 sont concernés. Du facteur 2, tous les items sont concernés. Aussi, 21 des 32 items ne sont pas retenus dans cette solution factorielle.

En contexte exploratoire, comme lors d'une démarche de validation d'un instrument, Durand (2003) utilise la stratégie d'assouplir les exigences des seuils de saturation jusqu'à 0,1. La solution factorielle obtenue, bien que faible, constitue une base pour travailler la validité conceptuelle de l'instrument, un point de départ. Le tableau 15 présente la solution factorielle obtenue selon cette dernière règle de décision. Elle implique 31 des 32 items. Un item B11 n'atteint pas ce seuil pour aucun des facteurs. Autre particularité, cinq items présentent une saturation à valeur négative (A03, A04, B11, B18, B20). Ils contribuent tous à la variance du 1^{er} facteur. Tabachnick et Fidel (2013) recommandent de considérer prudemment la contribution de ces items.

Tableau 15

Solution factorielle retenue lors de la réduction de dimensions des items du QuizCI.

| Items | Facteur 1 | |
|-------|-----------|--------|
| B07 | 0,791 | -0,098 |
| A12 | 0,756 | -0,053 |
| B13 | 0,623 | 0,044 |
| A10 | 0,577 | -0,066 |
| A03 | -0,559 | 0,197 |
| B20 | -0,540 | 0,165 |
| B10 | 0,525 | 0,095 |
| A11 | 0,516 | 0,106 |
| A04 | -0,478 | 0,282 |
| B15 | 0,474 | 0,018 |
| B18 | -0,394 | 0,250 |
| B14 | 0,364 | 0,057 |
| B12 | 0,244 | 0,093 |
| A08 | 0,107 | 0,027 |

(Tableau se continu)

| Items | | Facteur 2 |
|-------|--------|-----------|
| B09 | 0,133 | 0,361 |
| A07 | -0,085 | 0,357 |
| A01 | -0,218 | 0,325 |
| B02 | 0,027 | 0,324 |
| B16 | -0,133 | 0,319 |
| B08 | 0,011 | 0,302 |
| B06 | 0,073 | 0,267 |
| B04 | -0,115 | 0,218 |
| A05 | 0,083 | 0,205 |
| B01 | 0,020 | 0,203 |
| B19 | -0,056 | 0,201 |
| B05 | 0,068 | 0,192 |
| A06 | -0,115 | 0,154 |
| B17 | 0,067 | 0,125 |
| A09 | -0,015 | 0,118 |
| B03 | -0,076 | 0,116 |
| A02 | -0,035 | 0,101 |

Une analyse de la concordance entre la solution factorielle à deux facteurs permet tout de même de reconnaître la conceptualisation théorique de la CI. En effet, il est comme présenté dans le tableau 15, la solution factorielle correspond systématiquement à la distribution des items du QuizCI par niveaux cognitifs d'apprentissage inférieurs et supérieurs telle que définie dans le tableau de spécification validé (tableau 8). Le tableau 16 regroupe les items du premier facteur comme vérifiant l'acquisition et la compréhension des connaissances, alors que les items du second facteur représentent les niveaux cognitifs supérieurs : application, analyse, synthèse et évaluation. Quinze items saturent davantage sur le facteur 1; tandis que dix-sept saturent sur le facteur 2. Cette correspondance nous permet de nommer les facteurs « Facteur contribuant à la mesure de la connaissance et de la compréhension de la CI » et « Facteur contribuant à la mesure de l'acquisition de niveaux cognitifs supérieurs ».

Ces deux dimensions de la solution factorielle correspondent aux niveaux cognitifs inférieur et supérieur des items et non pas aux dimensions conceptuelles de la CI. Aussi, nous avons vérifié la cohérence interne des items pour chacun des facteurs. Or, le coefficient alpha

de Cronbach obtenu pour chacun est respectivement insatisfaisant lorsque calculé par facteur ($\alpha = 0,423; 0,401$), alors qu'il est satisfaisant ($\alpha = 0,636$) pour l'ensemble des items.

Tableau 16

Étiquettes de la solution factorielle des items du QuizCI.

| Items du facteur 1 contribuant à la mesure de la connaissance et de la compréhension de la CI | Items du facteur 2 contribuant à la mesure de la CI dans son application, ou en contexte d'analyse, ou de synthèse ou d'évaluation |
|--|---|
| A03, A04, A08, A10, A11, A12, B07, B10, B12, B13, B14, B15, B18, B20 | A01, A02, A05, A06, A07, A09, B01, B02, B03, B04, B05, B06, B08 B09, B16, B17, B19 |
| $\alpha_{f1} = 0,423$ | $\alpha_{f2} = 0,401$ |
| $\alpha_{\text{QuizCI}} = 0,636$ | |

De surcroît, conceptuellement, il est illogique de dissocier les dimensions de la CI. Il s'agit d'un continuum visant à satisfaire le besoin d'information. Ainsi, *Accéder aux ressources* ne prend son sens que lorsque l'étudiant montre qu'il peut *Évaluer l'information*, ou en faire un *Usage éthique*. L'étudiant ne peut être estimé compétent qu'à une de ces dimensions. Il maîtrise l'information lorsqu'il peut combler son besoin d'information. Aussi, nous considérons que les analyses ne fournissent aucun argument contredisant notre conceptualisation; au contraire, que les analyses fournissent de meilleurs résultats lorsque la CI est traitée comme un construit unidimensionnel.

Le résultat de l'analyse de la cohérence interne est sans doute biaisé par le nombre d'items à partir duquel est calculé le coefficient. Cependant, ce construit correspond à la mesure attendue, représente mieux notre conception de la CI et constitue un résultat qui va dans le même sens que le résultat de l'AFE. Donc, nous considérons notre instrument le QuizCI comme étant censé capter une mesure unidimensionnelle.

Ainsi, une analyse factorielle exploratoire a été réalisée afin d'examiner la correspondance entre la structure dimensionnelle attendue et celle observée. Les données du QuizCI se sont avérées factorisables et presque tous les items (31/32) contribuent à la solution factorielle obtenue. En outre, cette représentation correspond étroitement au construit théorique ayant permis l'élaboration de l'instrument, reflétant les niveaux cognitifs inférieurs et supérieurs des items.

4.4. Application du modèle de Rasch et informations sur la validité

Le troisième domaine de la validité conceptuelle examine les liens entre la CI des étudiants et leur score au QuizCI. Pour ce faire, les données collectées avec le QuizCI sont modélisées par l'application d'un modèle de Rasch. Le modèle estime pour chaque item la valeur de son paramètre de difficulté sur une échelle Logit en tenant compte du fait que les étudiants ont plus ou moins réussi cet item, par rapport aux autres items de l'épreuve. De la même façon, l'estimation de l'habileté des personnes prend en compte la difficulté des items. Ainsi, même si deux étudiants ont réussi et échoué le même nombre d'items, ils n'obtiennent pas forcément le même niveau d'habileté. L'habileté est en quelque sorte pondérée par la difficulté des items.

Le positionnement des paramètres d'items et de sujets sur une échelle unique fait que les écarts entre les items ou les sujets sont connus, ainsi que le rapport entre l'habileté des personnes et la difficulté des items. Une carte des personnes et des items, comme celle présentée à la figure 7 représente le positionnement des personnes et des items sur l'échelle Logit. Ainsi, à elle seule, la modélisation permet de mettre en évidence certaines caractéristiques des données. Ces caractéristiques contribuent à fournir des informations supplémentaires sur la validité conceptuelle de l'instrument et des données qu'il collecte. Les analyses permettant de fournir ces informations s'inscrivent dans la théorie de réponse à l'item (TRI). Pour obtenir cette information, il faut compléter le processus suivant : d'abord, les analyses de la TRI reposent sur des postulats à démontrer. Aussi, il nous faut vérifier au préalable comment les données collectées avec le QuizCI répondent à ces postulats. Ensuite, le modèle de Rasch est appliqué à l'aide du logiciel Winsteps et un processus d'ajustement des données au modèle est réalisé. Puis, suite à ces démarches, l'information fournie par Winsteps sur l'instrument, ses items ou les données collectées peut être appréciée en regard de la validité conceptuelle du QuizCI.

4.4.1. Postulats de la théorie de réponse à l'item

Les postulats liés à l'application d'un modèle de réponse à l'item concernent des thèmes déjà abordés : l'homogénéité de la variance, l'unidimensionnalité du trait mesuré; auxquels s'ajoute l'indépendance locale des items.

4.4.1.1. Homogénéité

Au 4.1.3., nous avons identifié deux groupes dont la moyenne est statistiquement plus basse que les autres et que nous pourrions donc considérer comme provenant d'une population différente des autres. Pour favoriser l'ajustement des données au modèle de Rasch, nous avons vérifié si le retrait des données issues de ces deux groupes ($n=52$ personnes) modifiait les résultats. Ces analyses étaient réalisables puisque le ratio de la taille de l'échantillon passe alors de 15 :1 à 13 : 1; ce qui demeure satisfaisant. Nous avons constaté que le retrait des patrons de réponses des étudiants de ces deux groupes est perceptible, mais qu'il n'affecte ni l'ordre ou le niveau de difficulté des items, ni l'habileté des autres étudiants. Nous avons donc choisi de ne pas tenir compte de cette différence de moyenne, de conserver notre échantillon de 469 sujets et de ne pas présenter ces résultats en détails.

4.4.1.2. Dimensionnalité

La dimensionnalité des données a déjà été investiguée avec une analyse factorielle exploratoire en 4.3.1. La solution factorielle exige d'outrepasser la règle de décision établie de ne retenir que les items saturant $> 0,3$. Winsteps propose également un dispositif pour apprécier la dimensionnalité du construit mesuré. L'analyse des vecteurs de dimensions principales captés par le QuizCI corrobore la solution factorielle retenue précédemment, soit la répartition des items sur les deux facteurs identifiés. Cependant, les items s'ajustent mieux au modèle de Rasch, tandis que la solution factorielle manque d'adéquation. Ce résultat révèle que l'objet d'étude la CI, telle qu'observée avec l'instrument QuizCI, présente une meilleure adéquation conceptuelle lorsqu'elle est examinée avec un modèle non linéaire comme le Modèle de Rasch, plutôt qu'avec un modèle linéaire comme l'analyse factorielle exploratoire. Aussi, nous estimons que la modélisation des données avec un modèle de Rasch constitue une avenue appropriée sur ce corpus de données, de plus que la cohérence interne de notre instrument est meilleure lorsqu'il est considéré unidimensionnel, ce qui concorde avec la conceptualisation théorique privilégiée.

4.4.1.3. Indépendance locale

L'indépendance locale est un postulat pour l'application du modèle de Rasch. En effet, la modélisation repose sur l'a priori théorique qu'une corrélation nulle est recherchée entre le score et l'erreur de mesure, afin que la probabilité de réussir l'item dépende uniquement du trait mesuré. L'indépendance locale des items contribue à la construction de la courbe de vraisemblance de réussite pour chaque individu en fonction des items. En ce sens, l'indépendance locale des items contribue à ce que pour deux individus ayant réussi le même nombre d'items, chacun n'obtient pas la même habileté (thêta), car la contribution de chaque item n'est pas équivalente, mais unique (Bond et Fox, 2007).

En contexte empirique, pour l'application du modèle de Rasch, les items doivent répondre au postulat d'une indépendance locale *suffisante*. Selon la théorie des probabilités, « la probabilité que se réalisent deux ou plusieurs événements aléatoires indépendants est égale au produit des probabilités associées à la réalisation de chacun de ces événements. » (Reischer, 2002). Bref, une trop grande corrélation inter-items suppose que le couple d'items est lié : soit qu'ils s'influencent et biaisent le résultat; soit qu'ils mesurent la même chose et que l'un d'eux est inutile à la mesure (Linacre, 2002).

Le logiciel Winsteps propose un examen des principaux coefficients de corrélation à partir des résidus standardisés (largest standadized residual correlation). En prémisses à une modélisation, Bond et Fox (2007) suggèrent de viser à obtenir des coefficients de corrélation inter-items très faibles, soit un seuil de corrélation plus petit que 0,5. Cependant, une corrélation inter-items jusqu'à 0,7 est tolérable et invite à la révision des items inter-corrélés afin de faire ressortir leur apport distinctif. Les données issues d'items entretenant une corrélation au-delà de 0,7 doivent être retirées de l'échantillon, car le couple d'items est trop lié.

Le résultat de l'analyse à l'aide des principaux coefficients de corrélation des résidus standardisés pour les items du QuizCI a permis d'identifier un seul couple d'items présentant un lien. A12 et B07 partagent une variance *faible* soit $\rho = 0,54$. Sans éliminer ces items, ils doivent être révisés. L'ensemble de l'échantillon répond donc *suffisamment* au postulat d'indépendance locale.

Ainsi, les trois postulats ne sont pas pleinement remplis par nos données. Certains des groupes composant l'échantillon présentent une moyenne différente de certains autres groupes. Cependant, ce paramètre ne semble pas affecter la modélisation et nous considérons adéquat de conserver les données issues de ces groupes dans l'échantillon modélisé. Également, l'analyse de la dimensionnalité captée par l'instrument ainsi que l'indépendance locale des items ne sont pas idéales selon les recommandations, mais paraissent suffisantes pour réaliser la modélisation de l'échantillon.

4.4.2. Processus d'ajustement des données au modèle de Rasch

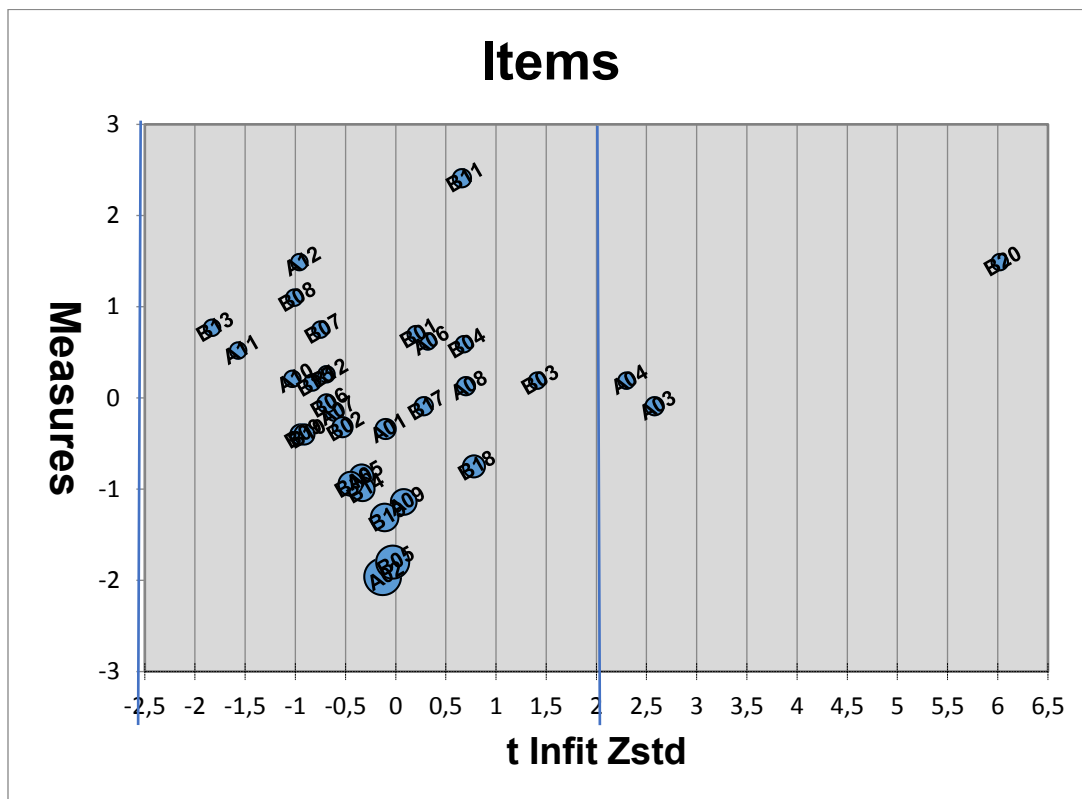
Les données collectées avec les 32 items du QuizCI auprès de 469 étudiants sont modélisées avec un modèle de Rasch appliqué à l'aide du logiciel Winsteps. Pour documenter l'ajustement des personnes et des items au modèle de Rasch, les coefficients Infit et Outfit sont les principaux outils utilisés. Tandis que pour apprécier l'ajustement, la carte des personnes et des items permet de visualiser le résultat de la modélisation.

4.4.2.1. Les coefficients Infit et Outfit

Les coefficients Infit et Outfit sont des coefficients mesurant l'adéquation des données au modèle après que les paramètres ont été estimés pour l'échantillon de données. Ils calculent la probabilité de la zone de résultat la plus vraisemblable, en raison d'une plus forte densité de données. Il existe 2 types de coefficients : un qui maximise l'écart entre les mesures en ciblant leurs dissemblances, le Outfit; un autre qui maximise les éléments communs, le Infit. Dans le cas du QuizCI, l'ordre et la position des items selon le Outfit et le Infit sont similaires, à la différence que le Outfit *étend* plus le modèle. Leur apport au processus d'ajustement est équivalent et fournit de l'information redondante. Aussi, seul le résultat de l'analyse du Infit est ici présenté.

Le graphique à bulles (figure 6) représente les items du QuizCI positionnés en fonction de la valeur de leur Infit (x) et de la valeur de leur mesure (y). Le Infit attendu selon Linacre (2002) doit se situer entre -2 et +2. Sur la figure, cet intervalle est marqué par des traits bleus et fait ressortir une problématique avec l'item B20 (Infit = 6,02). Les items A03 (Infit = 2,58) et A04 (Infit = 2,30) sont également problématiques, dans une moins grande proportion. Également, le

graphique à bulles permet de visualiser l'écart-type de l'item par la taille de la bulle. Il est notable que les items au bas du graphique, dont le niveau de difficulté est faible, présentent un plus grand écart-type, notamment A02 et B05.



Pour ajuster au mieux les données au modèle de Rasch, Linacre (2002) offre des lignes directrices :

1. Identifier les corrélations négatives des personnes ($\text{corr.} < 0$)
2. Vérifier l'ajustement des personnes à l'aide du Outfit carré moyen (cm); et du Infit carré moyen (cm) : $0,5 < \text{cm} < 1,5$.
3. Identifier les corrélations négatives des items : $\text{corr.} < 0$.
4. Vérifier l'ajustement des items à l'aide du Outfit carré moyen (cm) : $0,5 < \text{cm} < 1,5$.

S'il demeure des éléments problématiques, il suffit de réitérer ces étapes. Cependant, Linacre conclut que le chercheur doit faire preuve de jugement et ne pas retirer de données lorsque l'ajustement n'est pas amélioré par ce retrait (2002b). Le résultat de ces étapes peut être apprécié en regard de l'effectif d'items et de personnes, ainsi que par l'impact de ce processus sur les paramètres des items et des personnes.

4.4.2.1.1. Le résultat de l'ajustement de l'effectif d'items et de personnes

Ces vérifications ont permis d'identifier des items et des personnes problématiques à chaque étape :

1. Le score de 5 des 469 personnes (12022, 11029, 11019, 11023, 15011) présente une corrélation négative. Ces patrons de réponses doivent être retirés. Il reste 464 sujets.
2. Le Infit de quinze personnes ne se positionne pas dans les paramètres recommandés (12016, 10020, 12032, 12024, 06011, 16023, 10016, 03020, 05025, 10006, 16001, 10019, 14034, 16024, 11004¹²). Le retrait de ces patrons de réponses porte l'échantillon à 449 sujets.
3. Un item (B 20) présente une corrélation négative, soit plus petite que 0,0. Les données collectées avec l'item B20 doivent être retirées. Il reste 31 items.
4. Deux items (A03 et A04) présentent un coefficient Intfit trop élevé, soit supérieur à 2,0.
Le retrait de ces données porte l'effectif à 29 items.

Le processus d'ajustement maintient un ratio de la taille de l'échantillon à 15 patrons de réponses par items (15 : 1). Ce ratio est favorable à l'ajustement des données au modèle.

¹² pdelete = 288 461 426 307 411 303 269 438 256 118 20 266 460 134 341 313 349 270 333 339; idelete = 32, 3, 4

4.4.2.1.2. Les statistiques d'ajustement des données au modèle

Après la vérification des coefficients d'ajustement des données au modèle, nous consultons différents points d'information sur la qualité de la représentation fournis par le logiciel Winsteps. Les éléments les plus significatifs pour documenter la validité sont les indices de fiabilité et de séparation des personnes, les indices de fiabilité et de séparation des items. Ces informations évoluent lorsque des données sont retranchées. Ce sont ces informations qui, à l'aide d'une règle de décision, permettent d'apprécier les propriétés et les caractéristiques de l'instrument, de ses items ou des données collectées. L'évolution de ces paramètres est présentée, pour les personnes au tableau 17 et pour les items au tableau 18.

La séparation des personnes doit se distribuer convenablement le long de l'échelle (Linacre, 2002b). Pour cela, l'indice de séparation des personnes doit demeurer sous la valeur 2,0. Suite au retrait de 20 patrons de réponses, l'indice de séparation des personnes passe de 1,00 à 1,01; deux valeurs inférieures à 2,0, ce qui est adéquat. L'indice de fiabilité des personnes recommandé est supérieur à 0,80. Au cours de ce processus d'ajustement, il est passé de 0,50 à 0,51. Ces deux valeurs sont qualifiées de faibles. L'interprétation de Linacre du résultat (un indice de séparation fort du score des personnes, mais un indice de fiabilité faible) indique que l'instrument n'est pas suffisamment sensible pour distinguer les variations de compétence informationnelle des étudiants. Plus d'items peuvent être nécessaires (Linacre, 2002b).

L'indice de séparation et celui de fiabilité des items documentent l'adéquation de la hiérarchie des items. Les valeurs recherchées sont les mêmes que pour les personnes, soit : l'indice de séparation inférieur à 2,0; l'indice de fiabilité supérieur à 0,80. L'indice de séparation des items initialement est de 7,68. Il a diminué à 6,43 au retrait des personnes, puis remonté à 7,61 au retrait des items. Il est dans tous ces cas bien trop élevé. Linacre (2002) interprète ce résultat à l'effet que la hiérarchie des items n'est pas suffisamment perceptible dans l'échantillon de données ici modélisé. Ici encore se reflète donc le manque d'items dont le niveau d'habileté correspond au niveau de CI des étudiants. Linacre mentionne que cette problématique pourrait également être palliée par un très grand échantillon, soit $n > 2000$. L'indice de fiabilité des items lui n'a pas bougé au fil de l'ajustement et s'avère excellent à 0,98 (Linacre, 2002a). Ainsi, les items produisent une mesure constante.

Tableau 17

Effets du processus d'ajustement sur les paramètres des personnes

| Paramètres | Valeurs des paramètres avant l'ajustement | Valeurs des paramètres suite au retrait des personnes | Valeurs des paramètres suite au retrait des items |
|----------------------|---|--|--|
| Real person rmse | 0,45 | 0,45 | 0,45 |
| Model person rmse | 0,44 | 0,44 | 0,44 |
| Real sd | 0,43 | 0,44* | 0,44 |
| Model sd | 0,44 | 0,45* | 0,45 |
| Real separation | 0,95 | 0,97* | 0,97 |
| Model separation | 1,00 | 1,01* | 1,01 |
| Real p. reliability | 0,48 | 0,48 | 0,48 |
| Model p. reliability | 0,50 | 0,51* | 0,51 |

Légende : Les résultats en gras et étoilés précisent les résultats ayant variés suite à l'ajustement des données.

Tableau 18

Effets du processus d'ajustement sur les paramètres des items

| Paramètres | Valeurs des paramètres avant l'ajustement | Valeurs des paramètres suite au retrait des personnes | Valeurs des paramètres suite au retrait des items |
|-------------------|---|--|--|
| Real item rmse | 0,12 | 0,12 | 0,12 |
| Model item rmse | 0,12 | 0,12 | 0,12 |
| Real sd | 0,93 | 0,91* | 0,94* |
| Model sd | 0,93 | 0,91* | 0,94* |
| Real separation | 7,61 | 7,38* | 7,58* |
| Model separation | 7,68 | 7,43* | 7,61* |
| Real reliability | 0,98 | 0,98 | 0,98 |
| Model reliability | 0,98 | 0,98 | 0,98 |

Légende : Les résultats en gras et étoilés précisent les résultats ayant variés suite à l'ajustement des données.

À l'instar de Grondin et Dionne (2015, p. 14), nous constatons que le retrait des personnes ou des items identifiés problématiques fait assez peu fluctuer les statistiques d'ajustement, souvent seulement de quelques centièmes.

4.4.2.2. La carte des personnes et des items

La carte des personnes et des items (figure 7) représente l'habileté des personnes, ainsi que la difficulté estimée des items positionnés sur l'échelle en Logit. À gauche de l'échelle, les personnes sont représentées par # pour 6 étudiants ou un point pour 1 étudiant; à droite, les items sont représentés par leur numéro. Cette représentation peut être utilisée à deux fins. D'abord, elle contribue à identifier les données problématiques, particulièrement les items ou les personnes présentant des valeurs extrêmes. Pour ce faire, l'étendue est examinée. Puis, une fois l'ajustement des données au modèle finalisé, la carte des personnes et des items permet d'apprécier certains aspects de validité conceptuelle, soit la symétrie entre la difficulté des items et l'habileté des personnes, ainsi que la régularité de la distribution des items pour couvrir uniformément l'étendue de difficulté sur l'échelle dans l'intervalle prescrit.

4.4.2.2.1. L'étendue

L'étendue normale recherchée sur l'échelle en unités Logit, tant pour l'habileté des personnes pour que la difficulté des items, doit s'étendre de -3 à +3. Le zéro représente alors le point central (50% / 50%) de la probabilité de réussir ou d'échouer l'item. Les éléments qui se positionnent au-delà de l'étendue normale sont considérés comme extrêmes et nuisant à l'ajustement des données au modèle. Les sujets de l'échantillon collecté avec le QuizCI se positionnaient initialement sur l'échelle de -0,93 à +3,85. Trois personnes ont un score se positionnant sur l'échelle au-delà de +3,00 : personne₁₅₀₀₉ 3,85; personne₀₇₀₀₈ 3,08; personne₁₀₀₂₆ 3,08. Selon les recommandations de Linacre (2002), le patron de réponses de ces personnes doit être retiré de l'échantillon afin de favoriser l'ajustement des autres données au modèle. L'échantillon passe à 446 sujets. Les items du QuizCI, eux, s'étendent sur l'échelle Logit de -1,96 à +2,41. Cette étendue respecte la recommandation de Linacre (2002), soit de -3 à + 3 Logits. Aucun item n'est identifié comme aberrant parce que trop difficile ou trop facile.

Suite à l'ajustement des données au modèle, la carte des personnes et des items permet d'apprécier la modélisation finale des données en représentant l'étendue de la distribution des

personnes et des items, la symétrie de leur distribution, leur positionnement par rapport au zéro de l'échelle, ainsi que la régularité de leur distribution.

Dans notre modèle final, 29 items s'étendent sur l'échelle Logit de -1,96 à +2,41. Tandis que 446 sujets présentent une habileté CI de -0,93 Logit à +2,60 Logits.

4.4.2.2.2. La symétrie

La distribution des habiletés des sujets s'étend le long de l'échelle de +2,60 à -0,93. Les items se positionnent de +2,41 à -1,96. Ainsi, les items s'étendent d'environ 1 Logit plus bas sur l'échelle que les habiletés des personnes. De surcroît, visuellement, la carte permet de constater que la densité des scores totaux se situe en haut du zéro. Tandis que les items se positionnent majoritairement de +1,00 à -1,00. Cette asymétrie entre la distribution des scores des personnes et le positionnement des items sur l'échelle de mesure est marquée. Quinze sujets sur 446, soit seulement 3% de l'échantillon, ont un score total sous le zéro. Quinze items présentent une difficulté sous le zéro, tandis que 14 items se positionnent au-dessus du zéro. Un examen à partir du haut de l'échelle permet de constater l'asymétrie entre items et scores :

- Les quatre 4 items les plus difficiles (B11, A12, B20, B08) se positionnent dans le haut de l'échelle à un niveau de compétence estimé à +1,10 à +2,60. En face, 254 étudiants de l'échantillon ont une compétence estimée correspondant à ce même intervalle. Quatre (4) des 32 items (12,5%) permettent de représenter le niveau de CI de 254 des 469 sujets (54%).
- Seize (16) items¹³ du QuizCI, soit 50% des items, se positionnent dans l'intervalle central de +1,09 à +0,95 Logit; en face de 133 personnes (28% de l'échantillon).
- Tandis que les douze items les plus faciles (37,5% des items du QuizCI) se positionnent au bas de l'échelle dans l'intervalle de +0,95 à -0,93 Logit et représentent un niveau de CI trop facile pour 99% l'échantillon.

L'asymétrie entre l'habileté des personnes et la difficulté des items constitue une preuve supplémentaire à l'effet que l'épreuve est trop facile. Ce résultat est conforme à celui obtenu avec l'indice de simplicité (4.3.2.)

¹³ B13, B01, B07, A06, B04, A11, B12, A04, A10, B03, B15, A08, A03, A07, B06, B17

4.4.2.2.3. Ancrer le zéro

Le zéro de l'échelle constitue un point de référence. Un test équilibré devrait comprendre autant d'items au-dessus qu'au-dessous de ce point médian (Bond et Fox, 2007). Des items du QuizCI, 14 se positionnent au-dessus du zéro et 15 en dessous, ce qui constitue une distribution centrée. Aucun item ne présente un niveau de difficulté représentant précisément le zéro. L'item B06 est celui qui s'en rapproche le plus avec une valeur de -0,06 Logit (tableau 19).

4.4.2.2.4. La régularité de la distribution

Une autre caractéristique observable sur la carte des personnes et des items (figure 7) est l'irrégularité de la dispersion des items le long de l'échelle.

D'une part, certains intervalles ne sont pas couverts par les items du QuizCI. Les zones au-dessus de 1 Logit sont particulièrement peu représentées par les items du QuizCI. Il manque d'items pour rencontrer la CI des nombreux étudiants dont le niveau d'habileté se situe au-dessus de 1 Logit. Nous croyons que davantage d'items représentant l'intervalle entre +1,00 à +2,50 Logit permettraient de cerner plus finement la zone de compétence de nombreux étudiants. Tandis que 12 items, situés sous le zéro Logit, s'avèrent trop faciles pour l'échantillon. Presque toujours réussis par les étudiants, ces items contribuent à positionner la courbe de distribution des scores à l'échelle au-dessus du 50%. Il faudra discuter des implications de ce résultat.

D'autre part, la carte des personnes et des items (figure 7) pourrait laisser croire que certains items représentent un même niveau de compétence, par exemple les items A08, A10, B03 et B15. Toutefois, la consultation des valeurs de la mesure (tableau 19) des items permet de constater que seulement 2 items ont une valeur exactement identique B09 et B10. Le premier correspond à la composante *Accéder à l'information* du 3^e niveau cognitif *Application*; et le second, à *Utiliser l'information* du 1^{er} niveau cognitif *Connaissance*. Ainsi, les items B09 et B10 n'ont pas été élaborés de façon à se recouper. D'ailleurs, l'examen de l'inter-corrélation de ces items ne les a pas identifiés comme liés. Aussi, leur niveau de difficulté similaire n'est pas un problème.

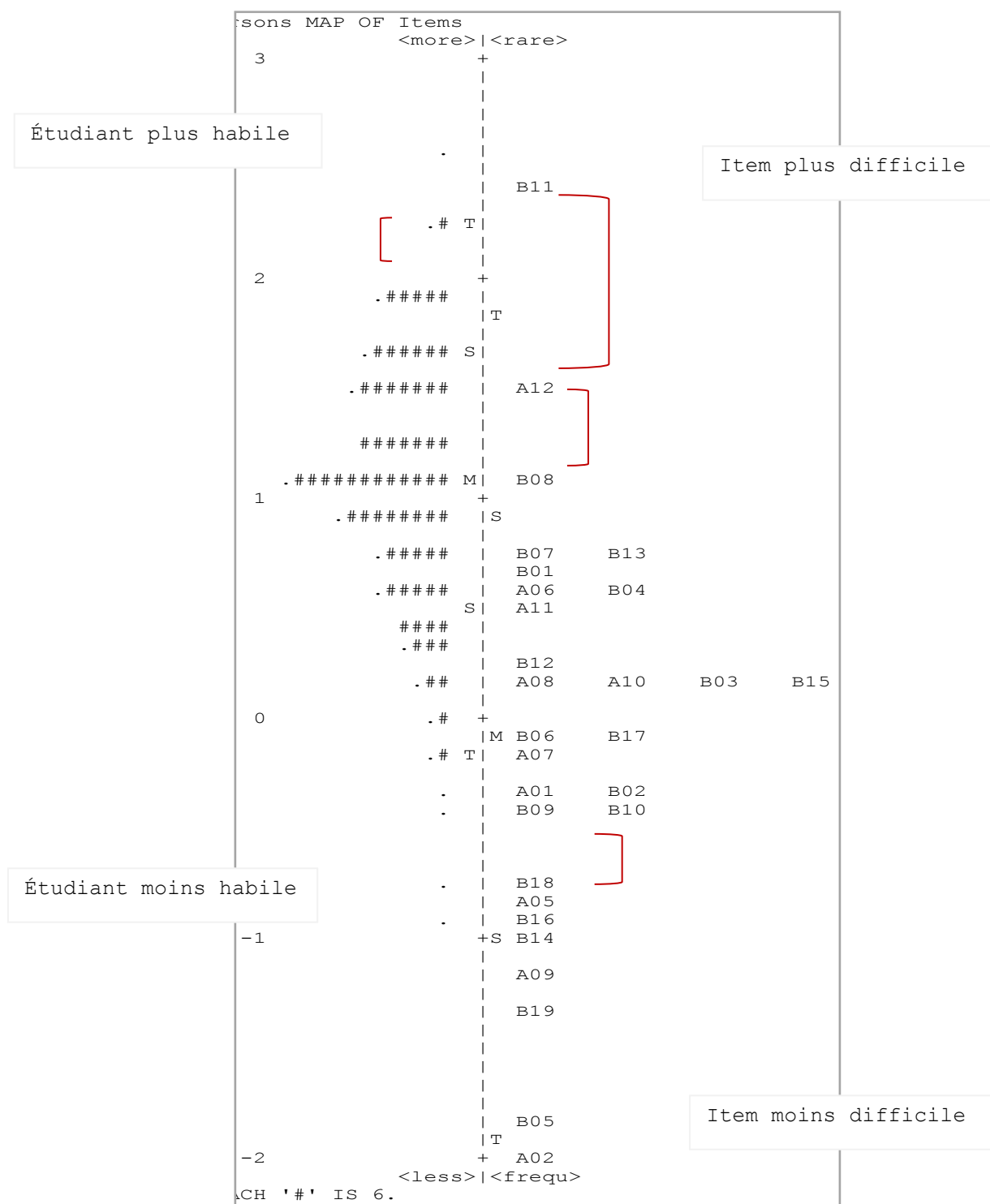


Figure 7. Carte des personnes et des items du QuizCI

Tableau 19

Mesures de la difficulté des items du QuizCI sur l'échelle de la CI

| Item | Mesure |
|-------------|--------------|
| B11 | 2,41 |
| A12 | 1,49 |
| B08 | 1,10 |
| B13 | 0,77 |
| B07 | 0,75 |
| B01 | 0,70 |
| A06 | 0,62 |
| B04 | 0,59 |
| A11 | 0,52 |
| B12 | 0,26 |
| A10 | 0,21 |
| B03 | 0,19 |
| B15 | 0,17 |
| A08 | 0,13 |
| B06 | -0,06 |
| B17 | -0,09 |
| A07 | -0,15 |
| B02 | -0,32 |
| A01 | -0,34 |
| B09* | -0,40 |
| B10* | -0,40 |
| B18 | -0,75 |
| A05 | -0,86 |
| B16 | -0,94 |
| B14 | -1,00 |
| A09 | -1,14 |
| B19 | -1,31 |
| B05 | -1,80 |
| A02 | -1,96 |

Légende : Les items en gras et étoilés précisent les items dont la valeur de la mesure est identique.

Finalement, malgré des postulats des méthodes de la TRI plus ou moins vérifiés, les données issues de 446 personnes et de 29 items se sont ajustées au modèle de Rasch. Celles de 23 personnes et de 3 items ont été retranchées parce que ceux-ci présentaient respectivement un coefficient Intfit trop élevé ou à corrélation négative. En outre, les statistiques d'ajustement au modèle fournissent des preuves supplémentaires du fait que l'instrument est trop facile pour les étudiants de l'échantillon :

- La séparation des personnes est adéquate alors que l'indice de fiabilité des personnes est faible. Cela suggère que l'instrument n'est pas suffisamment sensible pour distinguer les variations de compétence informationnelle des étudiants. Linacre (2002) suggère que plus d'items peuvent être nécessaires.
- L'indice de séparation des items est adéquat et celui de fiabilité des items bien trop élevé. Linacre (2002) estime que ce résultat montre que la hiérarchie des items n'est pas suffisamment perceptible dans l'échantillon de données ici modélisé. Ici encore se reflète que plus d'items correspondant au niveau de CI des étudiants serait nécessaire.

Cette asymétrie est clairement représentée dans la carte des personnes et des items obtenue, où : les items sont surreprésentés dans l'intervalle de -2 à 0 Logits, alors qu'ils sont trop peu présents l'intervalle entre +1,00 à +2,5 Logits. Cette représentation met en évidence le caractère irrégulier de la distribution des items selon leur niveau de CI.

Chapitre 5 : discussion sur la validité du QuizCI

Cette recherche a documenté la validité du QuizCI, un instrument de mesure de la compétence informationnelle (CI) des étudiants à l'Université du Québec. À cette fin, trois questions spécifiques de recherche ont été adressées :

Qs1 : Est-ce que le QuizCI réagit en fonction de la mesure captée?

Qs2 : Est-ce que le QuizCI représente la CI dans sa nature et sa structure conceptuelle?

Qs3 : Est-ce que les données collectées avec le QuizCI constituent une mesure de la CI?

Pour y répondre, nous avons produit de l'information principalement issue d'analyses quantitatives, dans le but d'en dégager des éléments de preuves de validité.

Lors de l'élaboration du QuizCI, des démarches pour documenter la validité apparente et de contenu ont été réalisées. Par la suite, nous avons constaté que les données collectées ont permis de réaliser les analyses voulues. Comme l'a mentionné Borsboom, Mellenbergh et van Heerden. (2004), il s'agit d'un prérequis pour documenter la validité conceptuelle. Une fois les données collectées, le chercheur doit composer avec l'échantillon en sa possession. Dans un cas où les données ne sont pas viables pour réaliser les analyses voulues, il faut en collecter de nouvelles. À cet effet, Laveault et Grégoire (2002), ainsi que Fox et Bond (2007) soulignent l'apport des dispositifs pour collecter des arguments de validité apparente et de contenu, de même qu'une rigoureuse connaissance et planification des analyses prévues pour documenter la validité conceptuelle.

5.1. Est-ce que le QuizCI réagit en fonction de la mesure captée?

Parmi la littérature recensée dans la présente étude, et dans la littérature en général (André et al., 2015), l'analyse fréquemment rencontrée pour vérifier si le test mesure bien ce qu'il est censé mesurer est l'analyse de la cohérence interne à l'aide de l'alpha de Cronbach. Ce dernier détermine la cohérence interne de l'instrument. Tandis que Bourque, Poulin et Cleaver (2006) rappellent que l'analyse factorielle exploratoire est la procédure la plus utilisée afin de déterminer si l'instrument « possède une structure et un contenu cohérents » avec le trait mesuré. Or, les informations produites par ces deux analyses au cours de la présente recherche nous ont permis de montrer que le QuizCI capte bel et bien un trait. Cette mesure d'une part présente une

cohérence interne qualifiée selon la règle de décision de Cronbach (1951) de satisfaisante ($\alpha_{\text{QuizCI}} = 0,636$).

Bond et Fox (2007) mentionnent que l'étendue de la distribution des données à l'intérieur des barèmes recommandés (-3 à +3), tant des personnes que des items; de même que la distribution des personnes sur l'échelle Logits en forme de cloche selon une distribution normale; informent que l'instrument prend une mesure et qu'elle est circonscrite. Les auteurs considèrent ces caractéristiques comme des éléments de preuve de validité.

Bref, l'ensemble de ces informations constitue des arguments en faveur de la validité du QuizCI à l'effet qu'il réagit en fonction de la compétence informationnelle telle que définie; ce qui répond positivement à la première question spécifique. Cependant, Laveault et Grégoire (2002) remarquent qu'il est rare qu'un instrument de mesure soit considéré validé avec une cohérence interne sous le seuil correspondant à excellent ($\alpha = 0,8$).

5.2. Est-ce que le QuizCI représente la CI dans sa nature et sa structure conceptuelle?

Bourque, Poulin et Cleaver. (2006) précisent : « La validité de construit implique qu'un instrument de mesure possède une structure et un contenu **cohérents avec ce qui est connu** du trait latent mesuré ». Nous avons adhéré à une conceptualisation du QuizCI unidimensionnelle, conforme à une continuité du processus requis pour combler un besoin d'information, compétence qui peut être intégrée par l'étudiant à différents niveaux cognitifs. Pour être compétent, l'étudiant maîtrise l'ensemble du processus en contextes complexes et nouveaux. Pour documenter la validité du QuizCI, lors de l'élaboration de l'instrument nous avons dans un premier temps porté une attention toute particulière à représenter les niveaux cognitifs inférieurs et supérieurs à l'aide d'un tableau de spécification (tableau 9, #2) que nous avons fait valider par quatre experts bibliothécaires (tableau 9, #8) et trois experts professeurs (tableau 9, #9).

Dans un deuxième temps, nous avons suivi la recommandation de Borsboom, Mellenbergh et van Heerden (2004) et avons fourni des preuves de la représentation conceptuelle à laquelle réfère la variable mesurée dans sa nature et sa structure conceptuelle grâce à une analyse factorielle exploratoire et une analyse des vecteurs de dimensions principales.

Nous avons relevé la faiblesse de la structure factorielle, alors que plusieurs items n'atteignent pas une saturation de 0,4. Cependant, en raison du contexte exploratoire de l'étude; d'une cohérence interne plus satisfaisante pour l'instrument que par sous-ensembles; ainsi qu'en raison de l'ajustement des données au modèle de Rasch, nous considérons ce résultat comme un point de départ valable et permettant d'identifier des pistes d'amélioration. Comme piste d'amélioration permettant de renforcer la structure factorielle, nous identifions l'élimination des items nuisibles à la cohérence interne de l'instrument et l'amélioration des items de difficulté insuffisante. Puis, rappelons que la taille de l'échantillon est sommairement suffisante. Enfin, Tabachnick et Fidell (2013) précisent que l'analyse factorielle permet une meilleure représentation avec un plus grand nombre de données

Ajoutons que nos résultats sont comparables à ceux de Timmers et Glas (2010), DeMars, Cameron et Erwin (2003); ou encore de Karim, Shah, Din, Ahmad et Khalid (2013). Lorsque ces trois études documentent la preuve de validité de leur instrument mesurant l'*Information Literacy*¹⁴, elles présentent des informations empiriques contribuant à la preuve que leur instrument mesure un objet unidimensionnel. Cependant, toutes trois obtiennent une preuve exploratoire. En effet, elle est respectivement obtenue en contexte local, auprès d'un assez petit échantillon. Aussi, ces résultats ne sont pas jusqu'ici généralisables parmi, par exemple, des populations d'étudiants universitaires de 1^{er} cycle (Reise, Waller et Comrey, 2000).

¹⁴ L'*Information Literacy* (IL) constitue le pendant anglophone de la compétence informationnelle (CI).

5.3. Est-ce que les données collectées avec le QuizCI constituent une mesure de la CI?

L'application du modèle de Rasch aux données collectées avec le QuizCI a fourni différentes informations en regard de la validité du QuizCI et de ses items. D'abord, le respect des trois postulats préalables à la modélisation n'est pas pleinement rempli par nos données. Cependant, Bond et Fox (2007), ainsi que Linacre (2002b) considèrent ce contexte tolérable lorsque les données s'ajustent au modèle. Ils identifient l'intérêt à poursuivre au niveau de l'information rendue disponible pour établir un premier constat sur la validité de l'instrument et un point de départ pour son amélioration. Le QuizCI s'inscrit dans ce processus. Il s'est effectivement ajusté au modèle de Rasch. Puis, l'information proposée par Winsteps nous permet d'identifier différents problèmes de validité du QuizCI. Ces informations obtenues, cumulées à celles fournies par les méthodes de la TCT, sont compilées au tableau 20. Leur considération comme un portrait d'ensemble permet de guider nos décisions pour l'amélioration du QuizCI.

5.4. Décisions pour l'amélioration de la contribution de l'item à la mesure

Pour l'amélioration du QuizCI, trois décisions sont possibles: conserver l'item tel quel, le réviser ou le retirer de l'épreuve.

Vingt items contribuent à la discrimination des étudiants en fonction de leur CI et à la cohérence interne de l'instrument. Huit sont d'une difficulté adéquate et douze sont faciles, mais tous se sont positionnés de façon symétrique avec l'échantillon de personnes sur l'échelle Logits. Puis, ils se sont ajustés au modèle de Rasch. En somme, les informations cumulées ne nous permettent pas d'identifier de problèmes de validité avec ces items. Aussi, ils seront conservés tels quels pour une prochaine collecte de données.

L'item B11 constitue un cas particulier. Il a été identifié comme nuisant à la cohérence interne de l'instrument, contribuant de façon insuffisante à la capacité de discrimination des étudiants en fonction de leur CI, en plus de ne pas avoir été retenu dans la solution factorielle. Cependant, lors de l'application du modèle de Rasch, l'item B11 s'ajuste au modèle. Nous expliquons ce résultat par le fait qu'il est le seul item à représenter un niveau de difficulté

supérieur à l'habileté des personnes. De surcroît, il se distingue du grand nombre d'items faciles pour les sujets de l'échantillon. Pourtant, selon les recommandations de Linacre (2002b), il ne représente pas un niveau de difficulté aberrant, soit +2,60 Logits, valeur qui est à l'intérieur de la recommandation de +3,00 Logits. Nous croyons donc qu'il faut réévaluer le rôle de B11 au sein du QuizCI lorsqu'il y aura davantage d'items pour représenter les niveaux de +1,00 à +2,50 Logits. À la lumière de ce résultat, nous suggérons de laisser l'item B11 tel quel pour des investigations ultérieures.

Ainsi, un total de 21 items (65% de l'effectif) sera conservé tel quel pour la prochaine collecte de données avec le QuizCI, dans le cadre des suites à donner à notre étude.

Le second cas de figure, réviser l'item, constitue le plus complexe puisqu'il incombe d'identifier en quel sens il devra être révisé.

L'item B03 est identifié comme nuisant à la cohérence interne du QuizCI et contribuant de façon insuffisante à la capacité de discrimination des étudiants en fonction de leur CI. Cependant, il se positionne de façon symétrique avec l'échantillon de personnes sur l'échelle Logits et s'ajuste au modèle de Rasch. Aussi, ces résultats nous amènent à conclure en faveur du fait que l'item B03 contribue à la mesure, mais qu'il doit être reformulé pour mieux travailler avec l'ensemble de l'instrument.

Un lot de six items contribue de façon insuffisante à la capacité de discrimination des étudiants en fonction de leur CI, car ils sont trop faciles pour les étudiants. Par ailleurs, ils présentent une cohérence interne satisfaisante et travaillent avec les autres items du QuizCI. Ils se positionnent de façon symétrique avec l'échantillon de personnes sur l'échelle Logits et s'ajustent au modèle de Rasch. Aussi, nous estimons que ces items doivent être révisés de façon à être plus difficiles. Cette décision pourra contribuer à rehausser la difficulté générale de l'épreuve, à améliorer l'ajustement général des données au modèle de Rasch, ainsi que les statistiques d'ajustement de l'instrument.

L'item B18 nuit à la cohérence interne du QuizCI et contribue de façon insuffisante à la capacité du QuizCI à discriminer les étudiants en fonction de leur CI. L'item s'ajuste au modèle

de Rasch, mais se positionne sur l'échelle Logits plus bas que le niveau d'habileté des personnes. Ainsi, une révision de l'item est requise, et ce, de façon à ce qu'il soit plus difficile.

Tandis que les items A03 et A04 nuisent à la cohérence interne du QuizCI et contribuent de façon insuffisante à la capacité de discrimination des étudiants en fonction de leur CI. Lors de l'application du modèle de Rasch, ces items présentent un Infit hors des barèmes recommandés, mais tout près. Pour être améliorés, ces items doivent être reformulés de façon à continuer de représenter adéquatement le concept de CI. Cependant, ils requièrent une révision en profondeur pour mieux contribuer à la mesure et concourir à rehausser la difficulté de l'épreuve.

Enfin, l'item B20 cumule à son actif uniquement des preuves de non validité. Il nuit à la cohérence interne du QuizCI. Il contribue de façon insuffisante à la capacité de l'instrument à discriminer les étudiants en fonction de leur CI. Puis, il ne s'ajuste pas au modèle de Rasch. Nous recommandons son retrait du QuizCI afin d'améliorer la validité de la mesure.

Tableau 20

Informations sur la validité de la mesure QuizCI et décisions pour l'amélioration du QuizCI

| Informations sur la validité | | | | | |
|---|---------------------------|---------------------|-----------|---|--|
| Décision | <u>Méthodes de la TCT</u> | | | <u>Méthodes de la TRI</u> | |
| | <i>D</i> | <i>A</i> | <i>IS</i> | <i>Symétrie</i> | <i>Ajustement au modèle de Rasch</i> |
| Conserver | | | | | |
| A06, A11, A12, B01, B04, B07, B08, B13 | Adéquat | Contribue | Adéquat | | |
| A07, A08, A10, B06, B12, B15, B17 A01, B02, B09, B10, B16, | Adéquat | Contribue | Facile | L'item Correspond à l'habileté des personnes | L'item s'ajuste au modèle |
| B11 | Insuffisant | α mieux sans | Difficile | | |
| Réviser | | | | | |
| B03 | Insuffisant | α mieux sans | Facile | | |
| A02, A05, A09, B05, B14, B19 | Insuffisant | Contribue | Facile | L'item se positionne sous l'habileté des personnes | |
| B18 | Insuffisant | α mieux sans | Facile | | |
| A04 | Insuffisant | α mieux sans | Facile | | L'item ne s'ajuste pas au modèle |
| A03 | Item nuit | α mieux sans | Facile | n.d. | |
| Rejeter | | | | | |
| B20 | Item nuit | α mieux sans | Adéquat | | |

Encore d'autres considérations doivent être tenues en compte pour l'amélioration de l'instrument.

Premièrement, il importe de conserver suffisamment d'items pour représenter les différents aspects de la CI et différents niveaux cognitifs, soit une possibilité de 24 nuances, bien que les items qui représentent les niveaux supérieurs puissent intégrer les quatre aspects du continuum de la CI.

Deuxièmement, l'asymétrie entre l'habileté des étudiants et la difficulté des items constitue le principal problème de validité du QuizCI. Cette asymétrie entre l'habileté et la difficulté nuit à la « hiérarchisation univoque et raisonnablement précise des personnes sur le continuum du trait mesuré » (Reise et al., 2000), soit la CI. Cette dernière considération doit être mise en perspective avec les arguments de validité apparente et de contenu décrits en 3.3.4. En effet, une fausse perception que le QuizCI est difficile est exprimée tant par les étudiants que par les bibliothécaires. Cette perception semble créer une certaine résistance à l'épreuve. Pour favoriser la validité de la mesure, tout en assurant l'acceptabilité sociale de l'épreuve, la modulation du score à l'épreuve pour la notation de l'étudiant pourra être considérée.

Autre considération, la carte des personnes et des items (voir figure 7) permet d'apprécier l'irrégularité des points de captation de la mesure en regard de la difficulté. La carte montre que certains niveaux de difficulté ne sont pas représentés par les items, ce qui constitue des intervalles non captés par l'instrument; tandis que d'autres niveaux sont représentés par plusieurs items. Or, lorsque la valeur de la mesure est prise en compte, seuls 2 items (voir tableau 19) représentent une difficulté exactement identique. Il se trouve que conceptuellement, ils ne représentent pas le même aspect de la CI (voir tableau 8). En raison de la complexité de l'étude de la régularité de la mesure, ainsi que la hiérarchisation des items décrite par Bond et Fox (2007), puis finement analysée avec le *Test Design Line* par Tristan et Vidal (Tristan et Vidal, 2006), cet aspect devra faire l'objet d'une étude ultérieure.

5.3. Limites de l'étude

Il reste difficile d'apprécier les résultats en comparaison avec ceux obtenus par d'autres études de validité. Les études recensées ayant utilisé une modélisation, soit celles de DeMars, Cameron et Erwin (2003) et d'O'Connor, Radcliff et Gedeon (2001), ne présentent pas leur résultat. Les auteurs expliquent simplement leur démarche. Récemment, Karim, Shah, Din, Ahmad et Khalid (2013) ont procédé à une analyse semblable. Ils ont cependant appliqué un Partial Credit Model, car leur instrument comprend des items à choix multiples dont le nombre de réponses possibles par item peut varier. Aussi, la démarche est semblable, mais le résultat difficilement comparable.

De plus, les résultats des analyses pour documenter la validité sont peu généralisables. D'abord, notre échantillon est composé d'étudiants de 1^{er} cycle à l'UQO Saint-Jérôme. Il ne peut pas représenter la population des étudiants universitaires, ou les étudiants de l'UQO en général. De plus, la taille de l'échantillon est relativement petite (Tabachnick et Fidell, 2013). Un plus grand échantillon permettrait plus de précision dans l'estimation des paramètres, autant pour l'analyse factorielle que lors de la modélisation des données avec le modèle de Rasch. Un plus grand échantillon renforcerait également le caractère généralisable des résultats.

Le choix du logiciel Winsteps, et non pas Rumm20-30, Bilog, Conquest ou R, influence le résultat par les fonctionnalités disponibles et en raison des algorithmes qui calculent les paramètres et les statistiques d'ajustement. Certains auteurs discutent de leurs caractéristiques et différences respectives de fonctionnement (Linacre, 2016) et d'estimation (Linacre, 2014; Sick, 2009).

Enfin, la durée limitée de l'étude laisse certains aspects sans réponses; notamment, la contribution des choix de réponse à la mesure par l'analyse de la courbe caractéristique de l'item (CCI) avec le logiciel Winsteps, ou encore, l'analyse des patrons de réponses qui ne se sont pas ajustés au modèle et l'erreur de mesure en général. Les sources possibles d'erreur, par exemple la tricherie, la motivation, les conditions d'administration (Penta, p.120) pourraient faire l'objet d'études ultérieures.

5.4. Prospectives

En effet, il beaucoup reste à faire pour disposer de données validées sur la compétence informationnelle des étudiants universitaires et avoir la possibilité de réaliser des inférences valides à partir des données collectées avec le QuizCI.

Afin de poursuivre le processus de validation du QuizCI, dans un premier temps il faut renforcer la preuve de validité en regard des trois domaines de la validité. Pour ce faire, il faut d'abord tenir compte des pistes d'amélioration du QuizCI identifiée dans la présente étude et modifier l'instrument en conséquence. Une analyse du contenu et de la formulation des items à la lumière des différents résultats de la présente étude est à faire. Puis, il nous faudra collecter des données avec ce *nouveau QuizCI* pour refaire les analyses permettant de documenter sa validité conceptuelle. Une telle démarche vise à obtenir une cohérence interne s'exprimant par un alpha de Cronbach supérieur à 0,80. Elle vise également à obtenir une preuve plus solide de la concordance entre la conceptualisation théorique et observée. Ensuite, certains résultats de la présente étude suggèrent que nos données s'ajustent mieux à un modèle non linéaire, ce qui suggère que les méthodes découlant du modèle de Rasch seraient mieux adaptées aux données collectées avec le *nouveau QuizCI*, que l'analyse factorielle exploratoire. Puis, suite à la modélisation des données collectées avec le *nouveau QuizCI*, nous viserons d'une part à obtenir un indice de séparation des personnes se distribuant davantage convenablement le long de l'échelle, d'autre part, un indice de fiabilité des items supérieur à 0,80 et un indice de séparation des items inférieur à 2,0. Selon les directives de Linacre (2002a), des items dont le niveau de difficulté correspond mieux à l'habileté des étudiants contribueront à atteindre ces deux derniers objectifs.

Enfin, dans le cas où les résultats aux démarches décrites ci-dessus sur les données collectées avec le *nouveau QuizCI* présenteraient des preuves de validité fortes et convaincantes, il serait alors intéressant d'examiner plus finement différents aspects : la courbe caractéristique des items (Bond et Fox, 2007); la régularité de la mesure, ainsi que la hiérarchisation des items à l'aide du Test Design Line (Tristan et Vidal, 2006) ; ou encore de collationner la mesure en des temps différents pour examiner la généralisabilité des données (Penta et al., 2005; Scallon,

1981); ou encore d'examiner les liens qu'elle pourrait entretenir avec d'autres variables (Salkind, 2010).

Conclusion

Au terme de la présente recherche, nous pouvons affirmer que le QuizCI présente un certain potentiel pour mesurer la CI de façon valide. L'instrument réagit en fonction de la CI. La mesure observée présente une cohérence interne. Le construit observé présente certaines correspondances avec la CI théoriquement énoncée. De plus, les données s'ajustent au modèle lors de l'application d'un modèle de Rasch. Également, cette étude a permis de mettre en lumière différents problèmes nuisant à la validité de l'instrument. Ces informations pourront contribuer à l'amélioration de l'instrument.

L'étude de la validité dans le cadre de la présente recherche s'est avérée exploratoire, du fait que peu de référents fournissent des pistes à suivre, et aucune en français. Également, le résultat reste peu comparable, encore une fois à cause du peu d'étude similaire disponible et aucune d'entre elles n'est généralisable. Aussi, cette étude constitue une pierre au bas du mur de l'étude de la mesure de la compétence informationnelle. Elle fournit un aperçu des possibilités méthodologiques pouvant être utilisées en sciences de l'information. Elle propose des pistes afin de produire des données valides, permettant de prendre en compte l'efficacité des actions et l'efficience des ressources investies en formation documentaire, ou démontrer la qualité des démarches entreprises par le bibliothécaire-formateur, et même favoriser l'imputabilité des acteurs universitaires. Ainsi, nous espérons contribuer à faire de la CI un objet dont l'étude est de plus en plus instrumentée.

Les sources documentaires

- Adam, T. et Maimets, I.-K. (2005). *Information Literacy Evaluation: Fishing for answers with SAILS*. Repéré le 2013-08-23 à <http://library.acadiau.ca/wilu/programme/SAILS.pdf>
- Allard, J. et Gérin-Lajoie, R. (2011). *CERTITUDE : vers une certification des compétences technologiques et informationnelles des étudiants accédant aux universités québécoises*. Montréal, QC: Maison des technologies de formation et d'apprentissage Roland-Giguère.
- Anderson, L. (1992). *Accroître l'efficacité des enseignants*. Paris, France: Unesco, Institut international de planification de l'éducation.
- André, N., Loye, N. et Laurencelle, L. (2015). La validité psychométrique : un regard global sur le concept centenaire, sa genèse, ses avatars. *Mesure et évaluation en éducation*, 37(3), 125-148.
- April, J. et Beaudoin, M. (2006). Projet d'intégration des compétences informationnelles: mise à l'essai d'un dispositif en enseignement préscolaire et primaire. *Documentation et bibliothèques*, 52(3), 173-181.
- Arias Coello, A. et Simón Martín, J. (2008). Student use of the library at the Complutense University of Madrid (Universidad Complutense de Madrid). *Estudio de los hábitos de uso de los estudiantes en la biblioteca de la Universidad Complutense de Madrid*, 31(3), 413-431.
- Association of College and Research Libraries. (2000). *Information Literacy Competency Standards for Higher Education B2*. Repéré le 2013-08-23 2013 à <http://search.ebscohost.com/login.aspx?direct=true&db=epref&AN=ILCSHE.ASSOCIATION.ACRL.BJJ.BJ&lang=fr&site=ehost-live>
- Big6 Team. (2012). *Big6 Skills Overview*. Repéré le 2013-08-15 2013 à <http://big6.com/pages/about/big6-skills-overview.php>
- Bond, T. G. et Fox, C. M. (2007). *Applying the Rasch model : fundamental measurement in the human sciences*. New Jersey, NY: Lawrence Erlbaum Associates Publishers.
- Borsboom, D., Mellenbergh, G. J. et van Heerden, J. (2004). The Concept of Validity. *Psychological Review*, 111(4), 1061-1071. doi: 10.1037/0033-295X.111.4.1061
- Bourgeois, S., Pouliot, J.-P. et Ruest, M.-E. (2011). *Intégration des compétences informationnelles*. Chicoutimi (QC): Université du Québec à Chicoutimi.
- Bourque, J., Poulin, N. et Cleaver, A. F. (2006). Évaluation de l'utilisation et de la présentation des résultats d'analyses factorielles et d'analyses en composantes principales en éducation. *Revue des sciences de l'éducation*, 32(2), 325-344. doi: 10.7202/014411ar
- Business-Higher Education Forum. (2003). *The Need for Changes in Teaching and Learning to Meet Global Challenges*. Communication présentée *Building a Nation of Learners*, U.S. Repéré à http://www.bhef.com/publications/2003_build_nation.pdf
- Catts, R. et Lau, J. (2008). *Vers des indicateurs de la maîtrise de l'information*. Paris, France: UNESCO.
- Centre national de ressources textuelles et lexicales. (2015). Ontologie. Dans Centre national de ressources textuelles et lexicales (dir.), *Le Trésor de la Langue Française Informatisé Ortholang*. France: Centre national de ressources textuelles et lexicales.
- Comité plurifacultaire d'éthique de la recherche. (2013). *La recherche nécessitant ou non une approbation éthique*. Repéré à <http://www.recherche.umontreal.ca/ethique-de-la-recherche/les-politiques-et-procedures/la-recherche-necessitant-ou-non-une-approbation-ethique/>

- Comité plurifacultaire d'éthique de la recherche. (2014). *Certificat éthique*. Repéré à <http://www.recherche.umontreal.ca/ethique-de-la-recherche/les-politiques-et-procedures/la-recherche-necessitant-ou-non-une-approbation-ethique/>
- Conférence des recteurs et des principaux des universités du Québec. (2005). *Norme sur les compétences informationnelles dans l'enseignement supérieur de l'Association of College & Research Libraries (ACRL)*. Repéré le 2013-08-23 à <http://www.crepuq.qc.ca/IMG/pdf/normeacrl-web-03-05-v4.pdf>
- Conférence des recteurs et des principaux des universités du Québec. (2008). *Tableau des compétences informationnelles et du niveau recommandé à l'entrée au 1er cycle universitaire CRÉPUQ*. Montréal: Conférence des recteurs et des principaux des universités du Québec.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Cureton, E. F. (1951). Validity. Dans E. F. Lindquist (dir.), *Educational measurement* (1st ed.^e éd., p. 621-694). Washington, DC: American Council on Education.
- DeMars, C. E. (June 2-5, 2002). *Modeling Student Outcomes in a General Education Course with Hierarchical Linear Models*. Communication présentée 42nd Forum for the Association for Institutional Research, Toronto (ON), Canada. Repéré à <http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED472657&lang=fr&site=ehost-live>
- DeMars, C. E., Cameron, L. et Erwin, T. D. (2003). Information Literacy as Foundational : Determining Competence. *JGE: The Journal of General Education*, 52(4), 253-265.
- Department of Education and Science (U.S.). (2002). *Partnership for 21st Century Skills*. Repéré le 2013-08-23 à <http://www.p21.org/>
- Direction des bibliothèques de l'Université de Montréal. (2006). *Banque de questions : Connaissances informationnelles et habitudes de recherche*. Montréal: Université de Montréal.
- Du Ruisseau, M., Gamache-Vaillancourt, G., Mallet, H., Roy, S. et Hiller, M. (2010). ING6900 Méthodes de recherche - Laboratoires. Dans École Polytechnique de Montréal (dir.), *Guide de l'étudiant*. Montréal: École Polytechnique de Montréal - Bibliothèque,.
- Dumont, R., Dupuis, M., Foucault, L., Hiller, M. et Proulx, M. (2005). *Innovative Approaches by École Polytechnique de Montréal Library in Support of Research Activities*. Communication présentée IATUL Annual Conference Proceedings. Repéré à <http://www.tandfonline.com/doi/pdf/10.1080/13614570500268316>
- Durand, C. (2003). *L'analyse factorielle et l'analyse de fidélité*. Montréal: Presses de l'Université de Montréal.
- Durnin, J.-F. et Fortier, C. (2008). *Appropriation du questionnaire d'auto-évaluation des connaissances informationnelles par les bibliothécaires : le cas des sciences de l'éducation et de l'anthropologie à l'Université de Montréal. (French)*. Communication présentée 74th IFLA General Conference and Council, Québec. Repéré à https://papyrus.bib.umontreal.ca/jspui/bitstream/1866/2587/1/134-Durnin_Fortier-fr.pdf
- Gagnon-Mountzouris, V. (2011, Août 2011). *Un test d'évaluation des compétences informationnelles à l'ÉTS*. Communication présentée Forum CI / IL, Montreal.

- Gervais, S. et Arseneault, C. (2005). Habiletés en recherche d'information des étudiants de première année universitaire en sciences de l'éducation. *Documentation et bibliothèques*, 51(4), 241-260.
- Grondin, J. et Dionne, É. (2015). *Les statistiques d'ajustement avec les modèle de Rasch*. Communication présentée ADMEE - Canada, Centre des congrès du Casino Lac Leamy - Gatineau.
- Groupe de Recherche sur la Culture et la Didactique de l'Information (2010). *Culture informationnelle et didactique de l'information: Synthèse des travaux du groupe de recherche 2007-2010*. France: GRCDI.
- Hinchliffe, L. J. (2005). Nouveaux développements en maîtrise de l'information. (French). *Bibliothèque nationale de France*, 50(6), 56-63.
- Hughes, H. (2010). International students' experiences of university libraries and librarians. *Australian Academic and Research Libraries*, 41(2), 77-89.
- IBM. (2010) *IBM SPSS Statistics 22*. Canada: International Business Machines,.
- International Federation of Library Associations. (2011). Recommandations de l'IFLA sur la Maîtrise de l'information et des médias. Dans International Federation of Library Associations and sInstitutions (dir.), *IFLA Headquarters Prins Willem-Alexanderhof*, (p. 2). Netherlands: IFLA Headquarters Prins Willem-Alexanderhof,.
- Karim, A. A., Shah, P. M., Din, R., Ahmad, M. et Khalid, F. (2013). Measuring information skills among Malaysian youth students: An instrument development. *Asian Social Science*, 9(16 SPL), 22-31. doi: 10.5539/ass.v9n16p22
- Katz, I. R. (2007a). ETS research finds college students fall short in demonstrating ICT literacy: National Policy Council to create national standards. *College & Research Libraries News*, 68(1), 35-37.
- Katz, I. R. (2007b). Testing information literacy in digital environments: ETS's iSkills assessment. *Information Technology & Libraries*, 26(3), 3-12.
- Kent State University. (2015). *About SAILS Project*. Repéré le 2013-08-23 à <https://www.projectsails.org/Background>
- Lakhal, S., Frenette, E. et Sévigny. (2013). Méthodes d'évaluation en administration des affaires: qu'en pensent les étudiants. *Mesure et évaluation en éducation*, 35(3), 117-144.
- Laveault, D. et Grégoire, J. (2002). *Introduction aux théories des tests en sciences humaines*. (2e^e éd.). Bruxelles: De Boeck.
- Lebrun, N., Perreault, D., Verreault, L., Morin, J., Raby, C. et Viola, S. (2007). Le développement des compétences informationnelles et son intégration disciplinaire dans un programme EPEP à l'heure des TIC. *Revue internationale des technologies en pédagogie*, 4(1), 47-55.
- Leech, N. L., Barrett, K. C. et Morgan, G. A. (2015). *IBM SPSS for intermediate statistics : use and interpretation*. (Fifth edition.° éd.). New York: Routledge.
- Lefevre, G. et Grenier-Gire, C. (2011). Panorama des recherches. Dans I. Fabre (dir.), *Professeur-documentaliste : un tiers-métier* (p. 87-110). Dijon, France: Éducagri éditions.
- Linacre. (2002a). Optimizing Rating Scale Category Effectiveness. *Journal of Applied Measurement*, 3(1), 85-106.
- Linacre. (2014). *Comparing estimates with other Rasch software*. Repéré à <http://www.winsteps.com/winman/compatibility.htm>

- Linacre. (2016). *Rasch Measurement Analysis Software Directory*. Repéré à <http://www.rasch.org/software.htm>
- Linacre, J. M. (2002b). Optimizing Rating Scale Category Effectiveness. *Journal of Applied Measurement*, 3(1), 85-106.
- Lonsdale, R. et Armstrong, C. (2006). The role of the university library in supporting information literacy in UK secondary schools. *Aslib Proceedings: New Information Perspectives*, 58(6), 553-569.
- McMillen, P. et Deitering, A. M. (2007). Complex questions, evolving answers: Creating a multidimensional assessment strategy to build support for the "Teaching library". *Public Services Quarterly*, 3(1-2), 57-82. doi: 10.1300/J295v03n01_04
- Megan, O. (2009). The information literacy instruction assessment cycle: A guide for increasing student learning and improving librarian instructional skills. *Journal of Documentation*, 65(4), 539-560.
- Mery, Y., Newby, J. et Peng, K. (2011). Assessing the reliability and validity of locally developed information literacy test items. *Reference Services Review*, 39(1), 98-122.
- Messik, S. (1989). Validity. Dans R. L. Linn (dir.), *Educational Measurement*, (3^e éd., p. 13-103). New York, NY: Macmillan.
- Microsoft. (2012). *LMsoftware*. Repéré le 2013-02-25 à <http://www.lmssoftinc.com/>
- Ministère de l'Éducation du Sport et du Loisir (2001). *Épreuves univformes du collégial*. Québec: Ministère de l'Éducation du Sport et du Loisir.
- Ministère de l'Éducation du Sport et du Loisir (2006). *Programme de l'école québécoise*. Québec: ministère de l'Éducation, du Sport et du Loisir.
- Mittermeyer, D. (2005). Incoming first year undergraduate students: How information literate are they? *Education for Information*, 23(4), 203-232.
- Mittermeyer, D. et Quirion, D. (2003). *Étude sur les connaissances en recherche documentaire des étudiants entrant au 1er cycle dans les universités québécoises*. Montréal, QC: Conférence des recteurs et des principaux des universités du Québec.
- O'Connor, L. G., Radcliff, C. J. et Gedeon, J. A. (2002). Applying Systems Design and Item Response Theory to the Problem of Measuring Information Literacy Skills. *College & Research Libraries*, 63(6), 528-543.
- O'Connor, L. G., Radcliff, C. J. et Gedeon, J. A. (2001). *Assessing Information Literacy Skills: Developing a Standardized Instrument for Institutional and Longitudinal Measurement*. Communication présentée Crossing the Divide: Proceedings of the Tenth National Conference of the Association of College and Research Libraries Chicago.
- Oakleaf, M., Millet, M. S. et Kraus, L. (2011). All Together Now: Getting Faculty, Administrators, and Staff Engaged in Information Literacy Assessment *Libraries and the Academy*, 11(3).
- Office national de la langue française. (2008). *Maîtrise de l'information*. Repéré le 2013-08-23 à http://www.granddictionnaire.com/btml/fra/r_motclef/index800_1.asp
- Organisation des nations unies (2003). *Vers une société compétente dans l'usage de l'information*. Prague: UNESCO.
- Organisation des nations unies (2013). *MIL : Global Media and Information Literacy Assessment Framework: Country Readiness and Competencies*. Japan: UNESCO Institute for Statistics.

- Panicker, S. (2005). Conference on "Minimal Risk" in Behavioral Science Research: A Decisional Framework for Investigators and IRBs. *Psychological Science Agenda*, 19(01).
- Penta, M., Arnould, C. et Decruynaere, C. (2005). *Développer et interpréter une échelle de mesure : Applications du modèle de Rasch*. Belgique: Pierre Mardaga Éditeur.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago, IL: University of Chicago Press.
- Reischer, C. (2002). *Théorie des probabilités : problèmes et solutions*. Sainte-Foy: Presses de l'Université du Québec.
- Reise, S. P., Waller, N. G. et Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, 12(3), 287-297.
- Réseau des répondantes et répondants en technologies de l'information et de la communication. (2011). *Profil TIC au collégial*. Repéré le 2013-05-06 à <http://reptic.qc.ca/voute/2011-10-profil-tic-grille-des-contenus/download.html>
- Salem, J. A. J. (2005). *Assessing Information Literacy Skills: A Look at Project SAILS*. US: Kent University.
- Salkind, N. J. (2010). Concomitant Variable *Encyclopedia of research design*. Thousand Oaks, CA: SAGE Publications Ltd.
- Scallon, G. (1981). La généralisabilité des seuils de performance établis par des appréciateurs : étude de trois procédés. *Revue des sciences de l'éducation*, 7(1), 3-21.
- Scharf, D., Elliot, N., Huey, H. A., Briller, V. et Joshi, K. (2007). Direct Assessment of Information Literacy using Writing Portfolios. *Journal of Academic Librarianship*, 33(4), 462-477. doi: 10.1016/j.acalib.2007.03.005
- Service de soutien à la formation - Université de Sherbrooke. (2011). *Littératie numérique*. Repéré le 2013-08-23 à <http://www.usherbrooke.ca/ssf/veille/numeros-precedents/septembre-2011/le-fin-mot-litteratie-numerique/>
- Shipman, D., Aloï, S. L. et Jones, E. A. (2003). Addressing Key Challenges in Higher Education Assessment. *The Journal of General Education*, 52(4), 335-346.
- Sick, J. (2009). Rasch Analysis Software Programs. *Testing & Evaluation SIG Newsletter*, 13(3), 13-16.
- Simard, S. (2009). An information literacy program built for relevance and purpose. *Reference Services Review*, 37(4), 386-394.
- Staley, S., Branch, N. A. et Hewitt, T. (2010). Standardised library instruction assessment: an institution-specific approach. *Information Research*, 15(3), 5-31.
- Tabachnick, B. G. et Fidell, L. S. (2013). Principal Component and Factro Analysis. Dans B. G. Tabachnick & L. S. Fidell (dir.), *Using multivariate statistics* (6^e éd., p. 612-680). Toronto, ON: Allyn and Bacon.
- Thirion, P. et Pochet, B. (2009). Information Literacy in Students Entering Higher Education in the French Speaking Community of Belgium: lessons learned from an evaluation. *IFLA Journal*, 35(2), 152-170.
- Timmers, C. F. et Glas, C. A. W. (2010). Developing scales for information-seeking behaviour. *Journal of Documentation*, 66(1), 46-69.
- Tristan, L. A. et Vidal, U. R. (2006, 4th July). *Linear Model to Assess the Scale's Validity of a Test*. Communication présentée New developments in Measurement Thinking, Chicago, IL.

- Université de Montréal. (2012). *Compétences informationnelles à l'Université de Montréal : Intégration au cursus*. Repéré le 2012-02-06 à <http://www.bib.umontreal.ca/ci/cursus/default.htm>
- Université du Québec. (2006). *Programme de développement des compétences informationnelles à l'Université du Québec*. Repéré le 2013-08-23 à http://www.uquebec.ca/~uss1109/dossiers/2005-2006/Journee-reseau_2006/depliant-COMINFO.pdf
- Université du Québec en Outaouais. (2009). *Outil de découverte*. Repéré à http://decouverte.uquebec.ca/primo_library/libweb/action/search.do?dscnt=0&dstmp=1397564909177&vid=UQO&fromLogin=true
- Université Laval. (2012). *Rechercher et présenter l'information*. Repéré le 2013-08-23 à <http://www.fse.ulaval.ca/reussir-etudes>
- Zubairi, A. M., Lide , N. et Kassim, A. (2006). Classical and Rasch analysis of dichotomously scored reading comprehension test items. *Malaysian Journal of ELT Research*, 3(2), 1-20.
- Zurkowski, P. G. (1974). *The information service environment relationships and priorities*. Washington: National Program For Library And Information Services, National Commission On Libraries And Information Science.

Annexe 1 : Évaluation de la formation par les étudiant



Évaluation de la formation aux compétences informatiques

Titre du cours : _____

Sujet du cours : _____

Été : ☐ Automne : ☐ Hiver : ☐

Nom de la personne ressource (bibliothécaire) : _____

Nombre de numéros à l'Université : _____

A. Habileté de la personne ressource

répondants

| Pour chacune des questions suivantes, veuillez cocher la case appropriée | Tout à fait d'accord ☹️ | Plutôt en accord ☹️ | Plutôt en désaccord ☹️ | Tout à fait en désaccord ☹️ |
|--|-------------------------|---------------------|------------------------|-----------------------------|
| A.1 La personne-ressource a fait preuve de structure et d'organisation | | | | |
| A.2 La personne-ressource s'est exprimée clairement (langage accessible) en se souciant de la compréhension des étudiants | | | | |
| A.3 La personne-ressource maîtrisait bien son contenu | | | | |
| A.4 La personne-ressource a utilisé des exemples qui ont facilité votre compréhension | | | | |
| A.5 L'animation de la personne-ressource a su maintenir votre attention et votre intérêt | | | | |
| A.6 La personne-ressource s'est montrée ouverte aux questions des étudiants (ex. suffisamment d'espace-temps pour les questions) | | | | |

B. Contenu de l'apprentissage

| | Tout à fait d'accord ☹️ | Plutôt en accord ☹️ | Plutôt en désaccord ☹️ | Tout à fait en désaccord ☹️ |
|--|-------------------------|---------------------|------------------------|-----------------------------|
| B.1 L'environnement physique a grandement facilité votre apprentissage (laboratoire, salle de classe) | | | | |
| B.2 L'utilisation d'outils technologiques a favorisé votre apprentissage (PowerPoint, démonstrations en ligne) | | | | |
| B.3 La documentation qui accompagnait la formation était claire et facile à comprendre | | | | |

C. Pertinence de la formation par rapport à votre domaine d'études

| | Tout à fait d'accord ☹️ | Plutôt en accord ☹️ | Plutôt en désaccord ☹️ | Tout à fait en désaccord ☹️ |
|--|-------------------------|---------------------|------------------------|-----------------------------|
| C.1 La formation est essentielle à votre programme d'études | | | | |
| C.2 La formation vous a permis de connaître les meilleurs outils de recherche | | | | |
| C.3 Vous comptez utiliser régulièrement les outils de recherche présentés (par exemple : pour améliorer la qualité de vos travaux) | | | | |
| C.4 Vous recommanderiez cette formation à d'autres étudiants | | | | |

... vers

Qui avez-vous trouvé le plus intéressant dans cette formation?



Qui avez-vous trouvé le plus difficile lors de cette formation et pourquoi?

Autres commentaires pour améliorer la formation :

Avez-vous déjà suivi une telle formation?

Non :

☐

Oui :

☐

Université :

Cégep :

☐